THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

PSYCHOLOGICAL ASSESSMENT OF MILITARY FEDERAL AGENTS

USING THE MMPI-2: A LOOK AT EMPLOYMENT SELECTION

AND PERFORMANCE PREDICTION

By

ANN P. FUNK

A Thesis submitted to the
Department of Psychology
in partial fulfillment of the
requirements for the degree of
Master of Science
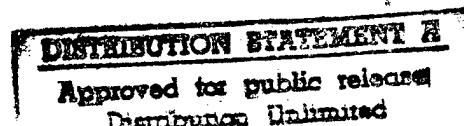
Degree Awarded:
Fall Semester, 1997

DTIC QUALITY INSPECTED 3

19971126 083

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE 19 Nov 97	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
PSYCHOLOGICAL ASSESSMENT OF MILITARY FEDERAL AGENTS USING THE MMP1-2: a LOOK AT EMPLOYMENT SELECTION AND PERFORMANCE PREDICTION

5. FUNDING NUMBERS

6. AUTHOR(S)
Ann P. Funk

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Florida State University

8. PERFORMING ORGANIZATION REPORT NUMBER

97-144

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES
83

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

DTIC QUALITY INSPECTED 3

The members of the Committee approve the thesis Ann P. Funk defended on October 30, 1997.
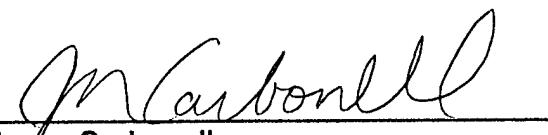
_____
Joyce Carbonell
Professor Directing Thesis

_____
John Brigham
Committee Member
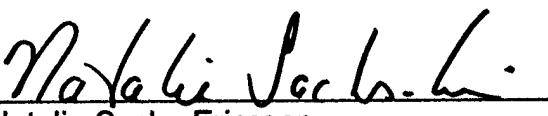
_____
Natalie Sachs-Ericsson
Committee Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study examined the utility of the Minnesota Multiphasic Personality

Inventory-2 (MMPI-2) for employment selection and performance prediction in a

sample of 133 military law enforcement investigators. An unscreened group who

completed the MMPI-2 but were hired without examination of the test results was

compared with a screened group whose results had been part of the hiring

decision.

In the first analysis, the MMPI-2 profiles of the two groups were

compared. Although the unscreened group tended towards higher mean scores

on the majority of scales, greater variability, and a higher frequency of elevated

scores, these trends did not reach statistical significance. Secondly, the two

groups were compared on three measures of job performance: a skills

composite score, number of positive distinctions, and number of negative

distinctions. The unscreened group tended towards greater negative

distinctions, although this trend likewise did not reach statistical significance. In

the final comparison, predictive links between the MMPI-2 protocols and job

performance criteria were explored using two subsets of the total sample. No

MMPI-2 scales were significantly linked to positive distinctions, five scales were

related to negative distinctions, and two scales were correlated with the skills composite score. However, the correlations were of insufficient magnitude to be used as individual predictors.

It seems most likely that previous non-psychological screening substantially reduced the magnitude of predictor-criterion relationships. Also, the small and unequal sample size limited statistical power. Nevertheless, this study described a unique sample of agents permitted to work without benefit of psychological test results--a methodological ideal. The results lend some support for the MMPI-2's usefulness, although further study is needed to more concretely address its utility in this population.

INTRODUCTION

Society has long recognized the need to carefully select police personnel. It is generally agreed that honest, intelligent, emotionally well-adjusted, and physically fit individuals are desirable in the stressful, unpredictable, autonomous, and "powerful" environment of police work. Although police agencies employed screening procedures such as character checks, medical exams, and personal interviews as early as 1829 in London, England, psychological screening of police applicants is far more recent. In the United States, police agencies began using psychiatric interviews in 1938. Around the 1950s, some departments began to use "modern" methods of objective and projective psychological assessment (Matarazzo, Allen, Saslow, & Wiens, 1964).

In 1967, the President's Commission on Law Enforcement recommended psychological screening of police applicants to ensure high professional standards and counter the abuse and unprofessional behavior demonstrated by some police personnel. Six years later, the National Advisory Commission on Criminal Justice Standards likewise recommended psychological examination as part of the selection process for law enforcement candidates (Scogin & Beutler, 1985). In 1977, the courts ruled that police agencies have a right to conduct

psychological evaluations. Four years later, they extended that right to an obligation, in effect, by ruling that police agencies may be held liable for the harmful or criminal behavior of employees whom they do not properly evaluate (Borum & Stock, 1993).

Usually, police departments ask psychologists to "screen out" applicants at high risk for job-related or public-safety problems. Psychologists use preemployment data to determine psychopathology or eliminate potential unsuccessful performers. Less frequently, psychologists attempt to "screen in" applicants, or use data to select potential successful performers (Fabricatore, Azen, Schoentgen, & 1978; Mills & Stratton, 1982). Similarly, the bulk of existing research examines the relationship between psychological variables and "screen out" decisions; empirical investigation of "screen in" decisions is sparse. Ideally, researchers would gather data on a sample of police applicants, permit them to work, and then relate performance criteria to previously developed data. Due to ethical concerns, this rigorous procedure of predictive validation has rarely been implemented. Instead, most studies on law enforcement selection employ concurrent validation, in which psychological variables of already-employed officers are examined to differentiate high and low performers (Bartol, 1991).

Regardless of the type of selection decision attempted, the recommended psychological screening battery for police applicants involves: gathering pertinent background or life history information, completing a personality inventory, assessing intellectual functioning, obtaining level of symptomatic

2

distress, and assessing interpersonal style (Bartol, 1991). The Minnesota Multiphasic Personality Inventory (MMPI) and its revised version, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) are the most frequently used assessment instruments in law enforcement screening (Hiatt & Hargrave, 1988a; Scogin & Beutler, 1985). In addition to its frequent use in personnel screening, the MMPI also appears to be the most widely used inventory for studying the effectiveness of screening (Johnson, 1983).

A review of the literature over the past couple of decades reveals the track record in the screening and selection of law enforcement personnel using the MMPI and MMPI-2 has been mixed. Some studies have reported minimal or nonexistent predictive power (Lester, Babcock, Cassisi, & Brunetta, 1980; Merian, Stefan, Schoenfeld, & Kobos, 1980; Mills & Stratton, 1982; Schoenfeld, Kobos, & Phinney, 1980). Other studies have reported some relationship between one, two, or more of the MMPI or MMPI-2 scales and some criterion of police performance (Bartol, 1991; Beutler, Storm, Kirkish, Scogin, & Gaines, 1985; Blau, 1994; Hiatt & Hargrave, 1988a; Shusman, Inwald, & Knatz, 1987).

In their critical review of police officer selection literature, Burbeck and Furnham (1985) stated:

> It should now be obvious that comparing the performance of
> successful and unsuccessful officers has not revealed any clear-
> cut differences that would be of use in the selection procedure.
> One problem seems to be that law enforcement in different types of
> police forces, different geographic areas, and different countries
> are not directly comparable, and it is not necessarily to be
> expected that one common denominator will be found (p. 64).

3

Whereas others (e.g., Blau, 1994) might disagree with this pronouncement about lack of clear-cut differences, many agree with Burbeck and Furnham's reasoning that study results are inconclusive in part because most of the research is atheoretical. Many researchers do not precisely analyze what police officers are required to do, and the skills necessary for those various tasks. When researchers choose tests based on organizational and job analysis, validity coefficients increase (Hogan, R, Hogan, J., & Roberts, 1996; Tett, Jackson, & Rothstein, 1991). Mills and Stratton (1982) echo the thought that assessment must be specific to an organization, because law enforcement organizations differ considerably in size, philosophy, and community services. To be fair, however, a minority of investigators argue "there is a stable, personological core to the notion of police effectiveness which transcends situational constraints" (Hogan & Kurtines, 1975, p. 294). Those who propose this viewpoint tend to be researchers whose work spanned earlier time periods.

Another reason why police officer selection research using the MMPI and MMPI-2 in particular is mixed is that many common pitfalls exist. According to Butcher, Graham, and Ben-Porath (1995), many researchers fail to completely describe local samples of convenience--the most common group--and overgeneralize from such samples to broader populations. They fail to have adequate comparison groups, improperly use normative data, and sometimes inappropriately match groups on variables other than the one of primary interest. Sample size is often too small to allow sufficient power, owing to the need to

correct for family-wise error, limited reliability of restricted-range criterion measures, and attenuated effect sizes typical of personality research. Hiatt and Hargrave (1988a) note that few studies report scale means and standard deviations for the criterion groups. Also, few studies compare the frequencies of elevated profiles between officers in different criterion groups, a potentially more useful statistic.

A further explanation for some of the variability in law enforcement selection research involves differences in what constitutes successful or unsuccessful performance, ideas of how to best measure it, and how much emphasis one should place on moderating variables. For example, the definition of successful performance (and unsuccessful performance at the opposite end of the spectrum) has ranged from favorable supervisory ratings (Schoenfeld, Kobos, & Phinney, 1980) to the absence of serious disciplinary actions (Hiatt & Hargrave, 1988a) to retention on the force for a minimum of three years (Saxe & Reiser, 1976). Moderating variables have included test taking defensiveness and motivation (Grossman, Haywood, Ostrov, Wasyliw, & Cavanaugh, Jr., 1990; Tett, Jackson, & Rothstein, 1991), gender and ethnic factors (Campos, 1989; Kornfeld, 1995), and degree of job autonomy (Barrick & Mount, 1993). They have also involved subject status (applicant, trainee, tenured employee), measurement method (external versus self-reported), and job complexity (Ones, Viswesvaran, & Schmidt, 1993).

<u>Current Study</u>

While it would be impossible in one study to capitalize on all of the lessons learned from prior research, as well as to consider all of the potential moderating variables, this study contributes to our understanding of law enforcement screening and selection in several unique and important ways. It involved a sample of military federal law enforcement investigators who completed the MMPI-2 as part of a preemployment battery, but were accepted for investigative duty without examination of the test results. They were compared with applicants accepted for duty under approximately equivalent conditions, but whose test profiles were screened for suitability prior to hiring. Of interest was whether the profiles of the unscreened group differed significantly from the profiles of the screened group. In addition, the job performance of the unscreened group was compared to the performance of the screened group. This predictive validation design is a research ideal rarely possible in police populations; it offered the potential to study the performance of officers with deviant profiles. To maximize the possibility of finding any differences which existed, a theoretical approach which took into account the federal agency's unique organizational culture, an analysis of personality and job skills required, and previous normative data was utilized.

To date, none of the research directly specified a sample of military agents, and studies of other federal agents are limited. The total sample is unique in this way. Further, the unscreened sample is unique in its own right,

quite apart from any comparison with other groups, because ethical considerations usually preclude the existence of such samples. In completely describing this sample, and reporting descriptive statistics for compared groups, this study also corrected some of the previously noted methodological problems.

Despite such measures, it was difficult to hypothesize about results. It seemed reasonable to predict that the unscreened group would show greater psychological profile variability than the screened group. But whether the groups would differ significantly from each other on job performance was unclear, given mixed findings in the literature as well as factors unique to this sample.

The performance criteria which were examined in this study--commendations, reprimands, supervisors' ratings, and unsatisfactory performance--have been shown by other researchers to significantly relate to MMPI profiles (Bartol, 1991; Beutler et al., 1985). These relationships held despite the fact that the applicants were previously screened by various measures such as background investigations and medical exams. Yet the preemployment measures of the law enforcement agency in this study were even more rigorous, and it was unknown whether the improved theoretical approach would overcome the presumed contribution of these findings.

Further, the base rate for screening out emotionally unsuitable applicants based on psychological test results might have been low. One large-scale study of an urban police agency showed that less than 3% of hired officers were

7

considered to be psychologically unsuited for law enforcement duties based on test results and a psychological interview (Shusman, Inwald, & Knatz, 1987). Another study of three police departments showed that fewer than 10% of applicants were screened out on the basis of MMPI profiles (Beutler et al.).

Finally, Hiatt and Hargrave (1988a) noted some problems in applying results reported in the literature to actual selection decisions. Often there is wide variability of scale means and standard deviations for criterion groups. Learning that a mean score differs by 3 points on a scale, between groups of officers, for example, is not useful when deciding about a specific applicant. This is particularly true because the lower level at which scale scores become predictive for law enforcement populations has not been systematically investigated. Accordingly, although the literature suggested some differences might be found on MMPI-2 profiles, it was unclear they would be large enough for practical significance (i.e., prediction of successful job performance).

LITERATURE REVIEW

Many issues arise when one uses personality measurements in preemployment screening and job performance prediction. R. Hogan, J. Hogan, and Roberts (1996) quite nicely outline major concerns involving: the nature of personality and its measurement, the uniqueness and commonality of various inventories, the appropriateness of personality inventories versus behavioral samples in predicting job performance, methodological flaws in research, and ethical and legal issues involving the use of personality measures (e.g., stability, validity, privacy, and discrimination). Each of these concerns has been extensively addressed in the general personality and personnel literature. Many of these issues have also been systematically explored in relation to the law enforcement profession. The results from the last two decades of research using the MMPI and MMPI-2 personality inventories will be presented.

MMPI

Butcher and Williams (1992) and Rogers (1995) report that Hathaway and McKinley began empirically developing the 550-item (later the 566-item), true/false, self-reported MMPI in 1939 to assist in assessing and diagnosing patients with mental disorders. It became the most widely used and researched

objective personality inventory in the world. In its earlier stages it contained four validity scales: Cannot Say (?), Lie (L), Infrequency (F)--sometimes called faking, and Correction (K)--sometimes called Defensiveness. It also contained ten clinical scales (alternately referred to by their name, number, or abbreviation): 1-Hypochondriasis (Hs), 2-Depression (D), 3-Hysteria (Hy), 4-Psychopathic Deviate (Pd), 5-Masculinity/Femininity (Mf), 6-Paranoia (Pa), 7-Psychasthenia (Pt), 8-Schizophrenia (Sc), 9-Hypomania (Ma), and 0-Social Introversion (Si). A scale elevated to a T score of 70 indicated clinical significance. Pertinent descriptions and behavioral correlates of these scales are given within the research summaries below and are also summarized in Appendix 1.

As time went on, researchers augmented the MMPI with countless supplementary scales, content scales, and indexes, to measure a diverse range of personality traits, behavioral predispositions, motivational factors, and symptomatic pictures. For example, there were scales involving ego strength, dominance, alcoholism potential, dissimulation, and neuroticism versus psychoticism. Researchers also developed subscales to clarify which of many possible heterogeneous symptoms comprised a particular elevation on a standard scale, as well as code types or clinical-scale summary indexes which described behavioral correlates for the most frequently occurring profile configurations.

Over time, researchers also established four common validity

configurations. One of them, the "inverted caret" occurs most frequently among defensive, "normal" individuals (such as unsophisticated job applicants). In this validity profile, the L and K scales are elevated above a T score of 60, and the F scale is near to or below a T score of 50 (Greene, 1991).

Despite its versatility, the MMPI had several problems, some of which developed over time. Many items were outdated or objectionable, the normative group was not representative of the population, the scales were unbalanced in proportion of true- and false-keyed items and overlapped considerably, and reliability and validity, though acceptable, were not as high as desired (Rogers, 1995).

MMPI-2

The MMPI was revised in 1989 to update its norms, revise outdated or nonworking items, and expand its measurement scope by adding new items and developing new scales. However, developers retained a large portion of the original instrument, including the traditional validity and clinical scales, to ensure continuity between the two versions and preserve a half-century of research. Psychometric properties, particularly for newly developed content scales, improved. The revised instrument, known as the MMPI-2, contains 567 items. A T score of 65 indicates clinical significance and falls uniformly across all scales at the 92nd percentile. This differs from the clinical T score of 70 on the MMPI which was thought to be at the 95th percentile but varied across scales (Butcher, Graham, & Ben-Porath, 1995; Butcher & Williams, 1992).

11

Equivalency

On the whole, research has shown the MMPI and MMPI-2 to be equivalent instruments (Butcher & Williams, 1992). Hargrave, Hiatt, Ogard, and Karr (1994) investigated whether this held true for a sample of 166 peace officers. They administered both versions to 96 men and 70 women. When grouping and comparing "normal," high-point, and 2-point codes, they obtained an overall concordance of 78%. A subset of "well-defined" profiles (those that had high points on two scales, the lower of which was at least five points higher than the remaining scales) produced a 90% agreement rate. Half of the subjects produced the same high-point code type, 33% of the men and 44% of the women gave the same 2-point code type, and 70% showed normal profiles on both tests. While initially appearing low, these percentages are similar to those produced by subjects who completed the original MMPI twice. All scales on both instruments were highly correlated, although two scales, 2 (D) and 5 (Mf), showed differences in raw score equivalence. For Scale 2, both genders scored lower on the MMPI-2. For Scale 5, men scored lower and women scored higher on the MMPI-2. Finally, the authors considered Scale K to be within normal limits, although T scores were 72 and 67 for the MMPI and MMPI-2, respectively. They based this on data from other studies showing law enforcement officers consistently score higher on this scale.

Hargrave et al. (1994) note that many of the officers in their sample were incumbents who had not completed the MMPI or MMPI-2 during their selection.

Thus, they may have differed from police applicants in that they may have had different motivational sets or work-induced characteristics. However, previous research has suggested comparability among the average profiles of officers and applicants (see discussion on participant status and defensiveness below).

Military Applicability

Military personnel (both men and women) were included in the MMPI-2 restandardization sample. In addition, Butcher, Jeffrey, Cayton, and Colligan (1990) explored the relevance of the MMPI-2 to a sample of 1,156 additional men aged 17 to 51 from the Army, Navy, Air Force, and Marine Corps. They found that special norms for military settings were not needed, as the service members obtained similar mean scores and factor structures as compared to the total restandardized sample.

Minority Representation

Much of the MMPI and MMPI-2 police research has involved predominately White male samples. A few studies address whether this instrument is equally valid for police populations of different ethnic and gender make up (new norms notwithstanding). Unfortunately, none of the research involves minority female police samples.

Campos (1989) investigated the psychological screening of Hispanic peace officers. He found that in 13 out of 16 MMPI studies, Hispanics scored significantly higher on the L scale than Anglos. There were also apparent Anglo-Hispanic differences on the clinical scales. However, Campos indicated

13

that any such differences were actually artifacts because they covaried with socioeconomic status and educational standing.

Muller and Bruno (1988) (as cited in Blau, 1994) administered the MMPI to 99 male police applicants divided into White, Hispanic, and Black triads and matched by age, education, and residence. Using the standard validity and clinical scales, they found no significant differences among ethnic groups.

Kornfeld (1995) gave the first 370 questions of the MMPI-2 as part of an employment screening assessment to 84 police applicants: 61 White males, 12 White females, and 11 minority males (representing three racial or ethnic groups). All groups showed defensive styles (elevated K scores), low scores on Scales 2 (D) and 0 (Si), and extreme scores on Scale 5 (Mf)--low for men, high for women. Whereas White male applicants differed statistically from minority male applicants on Scale 1 (Hs), no behavioral descriptors applied to the range of scores obtained (both were within normal range). There were no significant differences for any of the other scales. All of the clinical scale T scores were well below the cutting point of 65.

According to Keiller and Graham (1993), nonclinical men scoring low on Scale 2 are less likely to worry, to get hurt feelings, to have problems in making decisions, to give up easily, and to be overly sensitive to rejection. They are also more likely to be self-confident and to laugh and joke with people. Women scoring low on the same scale are less likely to worry, to complain of ailments, or to get nervous and jittery. They are also more likely to be cheerful. According to

14

Butcher and Williams (1992), men or women who score low on Scale 0 tend to be sociable, extroverted, friendly, active, interested in status, and competitive. They may act without considering the consequences and be opportunistic in relationships. Men who score low on Scale 5 endorse stereotypically masculine values, overemphasizing strength and physical prowess. They may be viewed as inflexible or coarse. Women who score high on the same scale describe interests typically seen as stereotypically masculine.

Although Kornfeld (1995) obtained an interesting police applicant "profile," neither he nor the other researchers derived significant differences among varying police populations. Thus, despite the lack of racial and ethnic diversification in many police officer samples, it would appear that screening of minority officers using the MMPI and MMPI-2 is not significantly different from screening of White-male officers.

Participant Status and Defensiveness

In addition to considering whether MMPI/MMPI-2 research is valid across gender, racial, and ethnic groupings of police personnel, some investigators have studied whether law enforcement psychological assessment results differ according to the status of participants: applicant, academy trainee, or tenured officer. This, in turn, leads to studies involving the applicability of MMPI/MMPI-2 norms to police populations and the extent of defensiveness or "faking good" in police test takers.

15

Gottesman (1975) compared the mean MMPI profiles of an experimental group of 203 urban North Jersey police applicants, a group of 89 Cincinnati police applicants, and a control group of 100 (police) veterans. He found that the mean profiles of police applicants were highly homogeneous and significantly deviated from the MMPI normative standardization group in consistent directions. He also found that the mean profile of the veteran controls was even more deviant from the normative group than were the applicant groups as compared to the normative group (excluding Scale K). Further, he noted that police applicant responses tended to be highly defensive, similar to the degree of "faking good" shown in other highly motivated groups in industrial screening settings. The author thus concluded that distinct personological variables and work needs might exist among urban applicants, MMPI norms may be inappropriate as a basis for assessing police applicant personality patterns, and some of the MMPI scale scores and related interpretations might be distorted.

Saxe and Reiser (1976) looked at 300 Los Angeles police applicants: 100 "successful" applicants (who passed a psychiatric evaluation and were still on the force), 100 "rejected" candidates (who failed the evaluation and were not appointed), and 100 "attrition" applicants (who passed the evaluation but separated from the force within three years). The authors found that all groups differed from the MMPI normative group on all scales except Hs. The applicants were lower than the normative group on the F and Si scales, higher on all remaining scales. The total applicant profile significantly differed from

16

Gottesman's New Jersey applicant profile on 8 of 13 scales. Both the L.A. sample and the N.J. samples differed from the MMPI normal profile in similar directions (excluding Pa which was lower than the normative group for the N.J. sample).

Saxe and Reiser further noted that successful applicants differed significantly from rejected ones on 6 out of 13 scales. The successful group was higher on L and K while the rejected group was higher on F, Hs, Pd, and Sc. The successful group also differed from the attrition group. The former were higher on L, K, Hy, and Pa while the latter were higher on Pt. Before trying to make too much of these differences or contrasting them with the findings of other studies, it is useful to know that the differences were all within the normal range and were too small in terms of traditional clinical standard scores to have meaningful utility. This study does underscore the potential dangers of using generalized test norms in selecting a specific vocational group, however.

In their meta-analytic review, Tett, Jackson, and Rothstein (1991) found that studies using recruits were significantly higher in validity than those using incumbents, a finding opposite to one prevailing theory that recruits would be more motivated to fake. Ones, Viswesvaran, and Schmidt (1993) found likewise. Though neither was limited to police populations, each revealed that presumed faking or presentation of favorable self-images among job applicants failed to have a significant depressing effect on validity measures.

Using a unique index and sample, Borum and Stock (1993) examined the differences in defensiveness between two groups of law enforcement applicants: 18 who admitted intentionally lying during the application process (deceptive group) and 18 for whom deception was neither admitted nor indicated (control group). The deceptive group scored significantly higher than the controls on Scales L and K and significantly lower (the expected direction) on a proposed new index, Es (Ego Strength) minus K.

The L scale is "a measure of the tendency of some individuals to distort their responses by claiming that they are excessively virtuous....[Elevated scores] suggest individuals who are presenting themselves in an overly positive light" (Butcher & Williams, 1992, p. 43). The K scale is "used both as an indicator of test defensiveness and as a correction for the tendency to deny problems....[It] is less 'obvious' in content than the L scale" (Butcher & Williams, p. 48). "Much of the research on Es suggests that it actually measures 'ability to withstand stress' more than potential for therapeutic success [its original purpose]" (Butcher & Williams, 1992, p. 171). "It measures physiological stability and good health, a strong sense of reality, feelings of personal adequacy and vitality, and spontaneity and intelligence" (Duckworth & Anderson, 1995, p. 324). Both Es and K "measure the effective operation of psychological defenses to bind psychological distress. Consequently, the [Es-K] comparison is made to differentiate the healthy defensiveness from the intentional effort to ignore or minimize difficulties" (Borum & Stock, 1993, p. 159). This comparison

18

had the best prediction rate of all variables studied, accurately classifying 83.3% of the deceptive applicants with a false positive rate of only 5.5%. Deceptive applicants scored significantly lower (the expected direction) on Es-K.

The groups in this study did not differ significantly on either the Obvious-Subtle total score or the F-K index. Research is equivocal about the utility of obvious and subtle scales in detecting defensiveness. Weed, Ben-Porath, and Butcher (1990) disproved the assumption that subtle scales are more valid than obvious scales because they are not subject to response bias. Comparing spouse ratings against self reports, there was a greater correlation with obvious scales. Hollrah, Schlottmann, Scott, and Brunetti (1995) also failed to find strong support for the validity of subtle items over obvious items (either as measures of their respective scale or as validity indexes). Yet some studies (including Grossman et al. below) have found otherwise. Research is likewise equivocal about the F-K index. Although designed to detect symptom exaggeration and defensiveness, it is much more successful in detecting the former rather than the latter. (Duckworth & Anderson, 1995).

Very few studies have assessed the utility of the MMPI validity scales when used with forensic samples. Even fewer have examined police officers involved in mandatory fitness-for-duty evaluations. Grossman et al. (1990) compared 40 Chicago police officers undergoing mandatory mental health evaluations for fitness to return to duty (20 strongly desiring to return--the positively motivated experimental group, and 20 specifically wishing not to--the

19

negatively motivated experimental group) with 20 officers not undergoing mandatory fitness evaluations (the control group). Four validity indices significantly differentiated experimentals from controls: Scales F, K, Ds (Gough Dissimulation Scale), and the Obvious-Subtle subscales. Officers wanting to return to duty minimized psychological difficulties significantly more than the other groups.

On scales the authors considered sensitive to both minimization and exaggeration (F-K and Obvious-Subtle), the positive motivation group had significantly lower scores (indicating they minimized more and exaggerated less) than the negative motivation group. The positive group also had lower scores than the negative group on two validity scales primarily sensitive to exaggeration (F and Ds). There were no significant differences between the positive and negative groups on two scales specifically responsive to minimizing (L and Mp). According to the authors, Mp--the Positive Malingering Scale--actually measures minimization rather than malingering. The control group's mean validity scale scores fell in between these two groups. For example, on the F-K Index (which yielded the highest percentage of clearly minimized profiles), 85% of the positives, 55% of the controls, and 45% of the negatives showed minimization (raw score difference of less than -11). On the Ds scale (which netted the highest percentage of clearly exaggerated profiles), 35% of the negatives, 10% of the controls, and 0% of the positives revealed some evidence of exaggeration (T score ≥ 61).

The study also found there was considerable minimization across all groups. In other words, both officers undergoing evaluations and a random-sample of active-duty officers showed response bias. Given this, Grossman et al. concluded that interpretations based solely on primary MMPI validity scales (L, F, and K) are insufficient for an adequate forensic evaluation of police officers and more liberal cutoff scores may be appropriate. Finally, the authors suggested that when response bias in the form of exaggeration does occur, it is mainly reflected on items assessing emotional or characterological problems (Obvious Depression, Hysteria, and Psychopathic Deviate subscales and the Ds scale), rather than items sensitive to major mental disorder. Thus, even a few symptoms suggesting psychosis may be more deviant for police personnel than the general population.

Taking the body of law enforcement research relating to participant status and defensiveness as a whole, a couple of conclusions seem justifiable. First, the MMPI profiles of police applicants, short-term employees, and veterans are relatively comparable. Although differences appear within these tenure groups and among study samples, the groups are more similar to each other than they are to the normative standardization group, from which they deviate in consistent directions. Second, law enforcement populations tend to minimize symptoms and show defensiveness (healthy or otherwise). Defensiveness as measured by the validity of subtle items over obvious items and the F-K index is equivocal.

Stronger support exists, however, when it is assessed via scores on the validity scales and (lower) overall elevations on the clinical scales.

Performance Prediction

Much of the MMPI law enforcement research involves attempting to differentiate psychological differences separating high and low, successful and unsuccessful performers. Approaches used have varied widely.

Matarazzo et al. (1964) reported on the characteristics of 116 successful police applicants and 127 successful firefighter applicants in Portland, Oregon. They were evaluated from 1959 to 1962 using various instruments; the MMPI was used from 1961 on. All had passed a Civil Service written exam, medical exam, physical agility test, and interview. Applicants considered "high risk" for failure were judged "clinically fragile" on MMPI profiles (n=84). Mean scores show successful policemen (n=35) and firemen (n=49) had remarkably similar profiles. Neither group contained any pathologically high scale scores; both elevated K, Pd, and Hy (T scores around 68, 65, and 55, respectively); and both were low on Si (T scores approximately 45). Extracting from one MMPI handbook, Matarazzo et al. interpreted these profiles as "typical of the enlisted men one often encounters in the military services: blustery, sociable, exhibitionistic, active, manipulating others to gain their own ends, opportunistic, unable to delay gratification, impulsive, and showing some tendencies toward overindulgence in sex and drinking" (p. 131). This characterization appears exaggerated given the limited elevations, not to mention stereotypical.

22

Merian et al. (1980) compared the MMPI protocols of 424 San Antonio policemen (obtained while they were cadets in training) against subsequent supervisory judgments of acceptability or unacceptability. Acceptability was decided by whether the raters would hire the officers again or would want them as backup in a crisis, given what they knew. Rather than using scales as most researchers do, the authors used individual items. They found 31 items which differentiated significantly between the unacceptable and acceptable officers, 5 of which replicated. They concluded, however, that "[While their results] appear to have sufficient validity to warrant further investigation,...the most parsimonious explanation of current results simply is that they are due to chance" (p. 158).

Mills and Stratton (1982) attempted to identify MMPI scores predicting success for Los Angeles county sheriffs at three levels: academy acceptance, academy graduation, and field employment. They used supervisory ratings based on easily observable, nonoverlapping job behaviors relating to characteristics such as energy level, self-confidence, aggression (altercations), organization under pressure, and decision making. A comparison of successful and nonsuccessful groups at all three levels showed no useful differences in MMPI scores. "Some comparative groups differed significantly on certain scales, [but] the strength of the relationships was very weak...[failing to] differentiate even the highest 10% from the lowest 10% of scores" (p. 13). One

drawback of this study, as published, is the authors' failure to report their methodology.

Beutler et al. (1985) obtained different results. They investigated the relationship between the results of multiple psychological assessment devices and in-service performance criteria of 65 officers in three police departments: one inner-city, one university, and one community college. They determined performance by supervisory ratings on two factors (interpersonal responsiveness and technical ability) and personnel data concerning reprimands, commendations, and grievances received; suspensions; referrals to counseling; and training attended. Four results related to the MMPI. Police officers judged high on technical ability tended to be somewhat depressed and suspicious (elevated scales 2 and 6). Those reprimanded for excessive use of force were low on indications of internal sensitivity (low Scale 1). Citizen grievances were related to officer depression levels (high Scale 2). Finally, departmental suspensions were strongly related to overall psychological distress (higher mean MMPI scale elevations).

An interesting feature of this study was that the researchers applied a principal-components analysis to numerous dependent ratings by supervisors to enhance statistical management. They also dichotomized infrequently occurring performance criteria to prevent a few officers with positive scores from skewing the results. That is, officers with a score of 1 or more in dimensions such as

commendations or grievances were contrasted with those who received zero scores (the majority).

As presented during a symposium in 1984, McCormick (as cited in Blau, 1994) also differentiated performance based on psychological testing. He attempted to predict effective job performance by comparing MMPI profiles of 60 "best" and 60 "least best" officers as rated by their watch commanders against 11 dysfunctional criteria. These included such things as excessive absenteeism or lateness; improper use of force or display of weapon; financial, alcohol, or drug problems; sexual harassment or petty thievery complaints; and deficiencies in report writing, knowledge of basic law, or court preparation. Higher elevations in Hs, Hy, Pd, and Ma scales (using a cutoff T score of 60) discriminated the "least best" from "best" group. McCormick replicated the study using 40 officers from another community, and reported an 80% hit rate.

Blau, Super, & Brady (1993) also replicated this study using 30 patrol, detention, and criminal investigative officers in a Florida sheriff's office. They found that the profile was equally effective in this jurisdiction, with hit rates for the three officer subtypes being 100%, 80%, and 60%, respectively. Though small numbers precluded further analysis, they hypothesized the lower hit rates for investigators resulted from the double screening process (detectives are usually hired from within the agency after proven performance).

Shusman, Inwald, and Knatz (1987) studied 698 male urban police officers in a probationary period using the MMPI and another measure obtained

25

prior to hiring. This group was previously screened in that any persons who had

negative test results had not been hired. Performance variables included

absenteeism, lateness, disciplinary actions, restricted duty, negative or positive

reports, and supervisory ratings. Multiple correlations between job performance

and discriminant scores ranged between .05 and .22 for the MMPI. The MMPI

correctly predicted "good" from "poor" job performance at a rate of between 43

and 75%. It improved hit rates over chance from 8% (final rating) to 38%

(lateness). Alcohol use (MAC scale), deviation from societal norms (Pd scale),

and defensiveness (K scale) weighed most heavily in the functions.

Hiatt and Hargrave (1988a) compared the MMPI profiles of 53 officers

from six departments involved in serious disciplinary actions against the profiles

of 53 officers not involved in such problems. Problem officers (47 men, 6

women) scored significantly higher on Scales F (Infrequency), 5 (Masculinity-

Femininity), 6 (Paranoia), and 9 (Hypomania) and significantly lower on Scale L

(Lie) than did nonproblem officers. Problem officers were twice as likely to have

a high-point elevation T score $\geq$ 70 as their nonproblem colleagues. With the

exception of Scales L and K (Correction), problem officers scored higher than

nonproblem controls on all scales. Hiatt and Hargrave concluded:

> These results indicate that any degree of psychopathology
> increases the likelihood of serious job performance problems. In
> addition, a presentation of self as conventional and moderately
> defended is associated with a lesser likelihood of job difficulty,
> whereas characteristics such as hypersensitivity, impulsivity, and
> poor frustration tolerance contribute to significant job problems (p.
> 722).

26

Of these eight studies, two found no differences between MMPI profiles and successful job performance. The remaining six had some overlap on the scales designated as predictive of poorer performance, as well as some uniqueness.

Longitudinal Studies

There have been very few attempts at predictive validation of the MMPI in police selection. In a longitudinal study, Azen, Snibbe, and Montgomery (1973) continued a study begun by Marsh in 1962 by following 95 Los Angeles county deputy sheriffs over 20 years using biographical, psychological, and aptitudinal variables. They found several MMPI scales to be predictive of successful performance. Scale 1 (Hypochondriasis) correlated to "rank status" or promotion (direction not specified). Higher scores on Scale 9 (Hypomania) and lower scores on Scale 2 (Depression) related to number of automobile accidents. However, the personality measures were obtained sometime after the individuals had been appointed, possibly contaminating their usefulness as predictive indicators.

Bartol (1991) followed 600 officers from 34 small-town police departments in Vermont over 13 years starting in 1975 to examine the ability of the MMPI to identify officers who would be fired due to poor performance. He ran Pearson correlations on the MMPI scores obtained before hire and average supervisory ratings of job performance as measured by behaviorally anchored rating scales (BARS). Bartol found none of the correlations were of sufficient magnitude to

27

individually use as predictive measures. He noted, however, that the higher the

Lie (L), Psychopathic Deviate (Pd), and Hypomania (Ma) score, the lower the

supervisory rating. In contrast, the higher the K or Hysteria (Hy) score, the

higher the rating. Using discriminant analysis, he developed an "Immaturity

Index" composed of Scales L, Pd, and Ma which correctly classified about 74%

of the officers, reducing the proportion of errors by 48%. When further

combining this Immaturity Index with department size and MMPI Scales K and

Hy, he correctly classified 80.04% of the officers, reducing the proportion of

errors by 60%. With the exception of the MacAndrew scale (MAC), none of the

supplemental MMPI scales contributed much. Even though the MAC scale

demonstrated significant correlations with many supervisory ratings and also

had a highly significant correlation with the Immaturity Index, it did not enhance

predictions.

One nice feature of this study was Bartol's attempts to ensure statistically

valid supervisory ratings. He spoke to chiefs and police administrators prior to

each evaluation process to explain his research, how to complete the rating

scales, and common rating errors. He also tested the reliability of the scales by

asking supervisors to immediately rate the performance of several officers whom

they knew well and to then rate them six months later. The retest correlation for

overall performance was .91. Another interesting aspect involved feedback from

police administrators. Bartol commented that,

> On the basis of over 15 years of feedback from police supervisors
> we consider an L score over 8 one of our best predictors of poor

performance in law enforcement. More recently, we have also discovered that extremely low L scale scores (0 or 1) also forecast poor performance, suggesting that the L scale may be curvilinear in its predictive power (p. 131).

Bartol also noted that the predictive power arising from his current model did not arise from scales resembling a classic 4-9 code. Rather, the average profile configuration of terminated officers showed moderate elevations (T scores between 50-55), with Scale 9 typically the highest.

These studies, similar to the eight shorter-term studies reviewed above, yielded mixed results. Two studies found individual scale scores to be significant predictors of job performance, while the third found significance only in using combinations of scales. Varying samples, dependent and independent measures, and methodology make comparison difficult, both for overall findings and the "consistency" of the predictive power of any individual scale. For example, scale 4 has been cited as predicting success in one study criterion, failure in another study criterion.

Clinician Accuracy

Considering that some researchers have supported the notion that the MMPI can predict important occupational outcomes, it is interesting to review studies which have compared the validity of clinicians' predictions to subsequent police performance. Schoenfeld, Kobos, and Phinney (1980) explored the interrater reliability of two experienced psychologists using the MMPI "clinically" in a simulated selection procedure. They reviewed the protocols of 424 San

29

Antonio policemen having 3 to 12 years experience, recommending acceptance or rejection. The authors correlated these recommendations with supervisory judgments of acceptability or unacceptability. This was the same sample used in the Merian et al. study, and acceptability was decided in the same manner (i.e., whether raters would rehire the officers or want them as backup in a crisis, given what they knew). The two psychologists used markedly different selection strategies and disagreed on recommendation for about one-third of the protocols. Neither was more accurate, however, when compared with supervisory ratings; nor did either perform significantly better than chance or classification by base rate.

The authors felt this "confirmed Levy's earlier warning that police applicants may be subjected to a variety of biases given lack of knowledge as to what constitutes emotional unsuitability or suitability for law enforcement work" (p. 424). Levy (1967) had two other pertinent cautions about screening approaches. First, the absence of unwanted qualities prior to employment does not guarantee a continued absence after hiring. Second, some traits often deemed pathological may be essential for the stress tolerance needed in effective police work.

Wright, Doerner, and Speir (1990) evaluated the relationship between preemployment psychological screening recommendations (based on MMPI, CPI, and interview results) and performance scores obtained by 131 Tallahassee police recruits during their initial field training program. This was a

4-phase program lasting 14 weeks, during which probationary officers received daily evaluations on five behaviorally anchored rating scales: appearance, attitude, knowledge, performance, and relationships. Each probationary officer was accompanied by a different senior training officer for each phase, and assumed more activities as competencies and phases progressed. Psychologist recommendations were unrelated to program evaluations during the first two phases and significant, though of modest magnitude, for the last two phases. No set of MMPI subscales explained or predicted performance scores from one phase to the next. However, Scale 2 (D) had a small negative correlation, and Scales 3 (Hy) and 4 (Pd) had small positive correlations with some BARS scores.

Hargrave (1985) obtained different results. He obtained MMPI profiles from 72 Los Angeles academy cadets on their first day of training. Using guidelines for interpreting the MMPI suggested by psychological skills analysis, three experienced clinicians labeled the profiles acceptable, unacceptable, or marginal. Hargrave then compared the clinicians' predictions to a rating of overall emotional suitability for law enforcement based on training attrition and instructor ratings. Significant agreement occurred in their decisions; using chi square analysis, the actual agreement was .70 with the expected chance agreement being .42. The author attributed the better reliability and validity to the linkage between MMPI interpretation and skills analysis data as well as to more specifically defined criterion measures. In fairness, however, note that

neither this study nor the one to follow compared MMPI profiles and police performance per se, but rather, academy performance. The two are not necessarily related.

In a similar study, Hargrave and Hiatt (1987) arranged to administer the MMPI to and psychologically interview 105 academy cadets who had not been psychologically screened prior to selection. As before, clinicians used a job analysis in their suitability decisions. Attributes which appeared most relevant fell into three broad groups: personality characteristics, interpersonal effectiveness, and intellectual characteristics. The psychologists also considered background factors in their decisions. Performance criteria involved training attrition, instructor ratings, and peer evaluations. Clinicians significantly agreed on suitability ratings whether based on test or interview results only or on combined results. However, test and interview predictions of overall suitability did not significantly correlate with each other. Further, clinicians ratings when compared to performance criteria differed based on method. Interview predictions were not statistically significant, test predictions approached significance, and combined predictions attained significance. This supports the view that different assessment procedures are complementary, each adding some unique dimension.

In a couple of rare instances, researchers were able to study the job performance of a group of incumbent officers who were hired by a department although judged unsuitable by evaluating psychologists. Hiatt and Hargrave

(1988b) compared 55 peace officers employed by an urban law enforcement agency; 40 suitable and 15 unsuitable, according to psychologists. According to supervisors, 31 were "satisfactory" (had no disciplinary actions and no more than one rating below satisfactory on any performance evaluation) and 24 "unsatisfactory" (had either numerous unsatisfactory ratings, were suspended or asked to resign, or had off-duty law violations). Evaluating psychologists had accurately classified 69%. Their main classification error was incorrectly designating 24% as suitable who subsequently performed unsatisfactorily. Only 7% (4 cops) found unsuitable by the psychologists received satisfactory ratings. Unsatisfactory officers scored significantly higher on Scales 6 (Paranoia) and 9 (Hypomania) than did satisfactory officers. Although not significant, the unsatisfactory group scored higher on 11 out of 13 scales. Scales K (Correction) and 3 (Hysteria) were the exceptions. Unsatisfactory officers had more qualities such as oversensitivity, rigidity, distrust, resentment, irritability, and maladaptive hyperactivity. Although this study had a small sample size, it provided some support for the validity of psychologists' decisions.

Lester et al. (1980) obtained a different result. In an initial study, they found a 76% congruence between a force psychologist's recommendations about 119 male police applicants and the department's decisions to hire or not hire. The department hired 43 of 50 (86%) recommended (fully) by the psychologist, 23 of 28 (82%) recommended with some reservations, and 17 of 41 (41%) recommended for rejection. In a second study, Lester et al. compared

the performance of 31 officers still on the force who had been in the "fully recommended" group with 15 officers still serving who had been in the "recommended for rejection" group. As judged by sergeants' and senior officer ratings, there were no significant performance differences.

Summarizing these six studies: two found no significant links between clinicians' predictions and "street" performance, three yielded small positive correlations between clinicians' recommendations and academy or probationary performance, and one found significant links between psychologists' appraisals and street performance, particularly for officer profiles designated as unsuitable.

Indirect Evidence

One other research effort indirectly relates to the effectiveness of MMPI personality factors in predicting police performance, as well as differences in indicators of success for training versus job performance. In a meta-analysis, Barrick and Mount (1991) studied the relation of the "Big Five" personality dimensions to job and training proficiency across five occupational groups. Police were one of the five occupations and constituted 13% of the samples. The big five refers to Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. "Many personality researchers now agree that the existing personality inventories [including the MMPI and MMPI-2] all measure essentially the same five broad dimensions with varying degrees of efficiency" (Hogan, R., Hogan, J., & Roberts, 1996, p. 470).

Barrick and Mount found Conscientiousness validly predicted successful

performance for all occupational groups. Openness to Experience and

Extraversion validly predicted training proficiency for all occupations. Most of

the correlations for Emotional Stability (Neuroticism) were relatively low. That

greater Conscientiousness predicted higher performance and Openness

reflected training readiness seemed logical. The other findings, though less

obvious, were likewise logical. The authors noted that most of the training

programs in these studies (such as police academy training) were interactive

and required a high energy level among participants. So higher correlations

between training success and Extraversion were not surprising. Barrick and

Mount explained the low Emotional Stability correlations as possibly due to a

"selecting out" of high neurotics from the labor force. Alternatively, they

proposed that as long as a person had "enough" stability, the predictive value of

any performance differences was diminished. The authors also suggested that

this finding explains why the overall validity of personality measures is relatively

low. Most scales such as the MMPI measure Emotional Stability, and none

measure Conscientiousness directly.

<u>Conclusion</u>

A review of studies conducted over three decades reveals psychologists'

have had mixed results when using the MMPI and MMPI-2 to screen and select

law enforcement personnel. Results varied depending on: year conducted

(1960-1970s, 1980-1990s); status of participants (applicant, trainee,

35

probationary officer, tenured officer); sample used (numbers, location, organizational culture); methodology employed (definitions used, measurement methods, statistical evaluation); type of study (concurrent or predictive validity; short-term or longitudinal); and focus of study (clinician accuracy, police performance). Not only did results differ according to these variables, they also differed when holding these variables constant. That is, largely similar studies had largely dissimilar findings. Factors which did not appear to appreciably alter results were individual race or ethnicity, and military or civilian status.

Nevertheless, some overall similarities were perceptible. First, earlier studies tended to have less conclusive results than later ones. Second, researchers using a theoretical approach tended to better predict occupational suitability or success, despite police applicants' or officers' susceptibility to defensiveness and other moderating variables. Third, with all due respect to empirical exceptions, common profiles emerged among police personnel. A large percentage of profiles (70% or greater) revealed MMPI scores in the normal range. Within this range, successful police officers tended to have moderate L and high K scales, as well as lower overall distress (mean scores on clinical scales). In contrast, unsuccessful police officers tended to have either very low or very high L scores; high F, 1 (Hs), 9 (Ma), and MAC scales; and higher overall distress. Both groups reported elevations on Scales 3 (Hy), 4 (Pd) and 6 (Pa).

This review revealed that the MMPI--and by virtue of its relative equivalency--the MMPI-2 held some promise for enhancing police employment decisions. Whether the MMPI-2 could merely raise the index of suspicion for detecting unsuitable candidates or could more meaningful predict successful performers with the sample herein remained to be seen.

METHOD

<u>Participants</u>

The participants were 133 military special agents who were selected (hired) by a federal law enforcement agency for investigative duty at worldwide locations between the years 1992 and 1996. This number represented all active duty (full time) military agents, both officer and enlisted, whose personnel and psychological records were intact and who were still on the job at the federal agency. Reserve (part time) military agents were excluded from this sample because they underwent different selection and training processes and had less time on the job from which to rate performance as compared to active duty agents. Civilian agents were also excluded because they underwent different selection processes than did the active duty agents.

The total sample was first divided into two comparison groups: an "unscreened" group of 116 agents and a "screened" group of 17 agents. This disparity in numbers existed because this study capitalized on a unique sample of convenience.

The unscreened group included persons who completed the MMPI-2 as part of a preemployment battery, but were accepted for investigative duty without

examination of the test results. Test results were normally analyzed by anagency clinical psychologist, but higher priority mission requirements prohibited timely screening in a group of applicants hired from late 1992 through early 1996.

The screened group included persons who completed an identical battery, but whose MMPI-2 test profiles were screened for suitability prior to hiring. These applicants were hired during late 1995 and early 1996. There were no applicants hired in 1992 through 1994 who fell in the screened group. It was not possible to obtain screened applicants prior to 1992 (owing to a change in archival methods) or from mid to late 1996 (owing to insufficient length of observation).

Demographic data relating to age, sex, race/ethnicity, education level, status (officer or enlisted), and years of experience for the two groups was collected. The total sample ranged in age from 23 to 39; the average was 29.56 years. There were 119 (89%) male agents and 14 (11%) female agents. Racial/ethnic composition included 112 (84%) Caucasians, 15 (11%) African Americans, and 6 (5%) "Other" minorities (Asians, Hispanics, and undifferentiated "other"). The majority had at least some college education; schooling ranged from 12 to 18 years with the average being 14.34 years. About three quarters of the participants were enlisted members; one quarter were officers. Staff Sergeant was the predominant rank. Participants occupied positions ranging in pay from approximately $22,860 to $45,760 annually.

Experience varied between one to four years with the average being 24.77 months.

There were no significant differences in the means for the two groups with respect to sex, race/ethnicity, education, and rank. With respect to age, the screened group was slightly younger ($\underline{M}$ = 28.12, $\underline{SD}$ = 3.06) than the unscreened group ($\underline{M}$ = 29.85, $\underline{SD}$ = 2.83), $\underline{t}$ (101) = -2.26, $\underline{p}$ = .025. Although age can affect MMPI-2 and performance scores, this difference of about one year and nine months is too small for practical importance.

Concerning experience, the screened group had fewer months of experience ($\underline{M}$ = 12.76, $\underline{SD}$ = 2.33) than the unscreened group ($\underline{M}$ = 26.58, $\underline{SD}$ = 8.92), $\underline{t}$ (96.88) = -13.65, $\underline{p}$ < .0001. This difference in experience was expected, given that the unscreened group was hired earlier than the screened group. As experience level could alter performance ratings, two additional comparison groups were created wherein subsets of screened and unscreened applicants more nearly matched on months of experience (these are described in procedures section).

Prior to hiring, all applicants had undergone a vigorous screening process involving records reviews (academic, job performance, training, medical, mental health, financial, and criminal); personal interviews and interviews of significant others (supervisors, coworkers, spouse, and neighbors); and various other procedures (sample of writing, achievement test for those without a college education, etc.) The process was aimed at determining an applicant's physical,

40

emotional, intellectual, and interpersonal suitability (and competitiveness) for agent duty, and occurred at multiple organizational levels. To become an agent, an applicant must first have received endorsement at the field unit, then passed screening at an intermediate level, and finally have been selected by a central board at the organization's headquarters. Typically, about 20-35% of enlisted applicants and about 1-5% of the officer applicant pool are selected for investigative duty each year (exact figures for the year groups in this study were not available).

Additionally, all agents completed psychological testing (including the MMPI-2) administered by agents in the field during the initial screening phase. For the unscreened group, the results were not used during the selection process. For the screened group, test results were presented to the central review board for consideration in context with other information. Findings would have been presented in one of three ways: "without indications of significant psychological concern," "needing clarification" (with issues of concern explained), or "indicative of potential problems."

Measures and Procedure

MMPI-2 scores for all participants were obtained from archival data. The 567-item personality inventories were individually administered during initial screening by agents in the field using standardized instructions. Protocols were scored at agency headquarters using a National Computer Systems program. Five of the raw scores were adjusted by adding a correction, based on the K

score, to compensate for test defensiveness, per custom, and all scale scores were converted into T scores using adult norms (Butcher & Williams, 1992). Since data was obtained via computer disk, a random sampling of protocols was manually replotted on MMPI-2 profile sheets to ensure data integrity (e.g., raw scores matched reported T scores, female scores on scale 5 had a different T score than identical male scores on scale 5, etc.) All protocols were evaluated for nonresponsive or random response patterns (i.e., Cannot Say raw score $\geq$ 30, F scale T score $\geq$ 110, True Response Inconsistency T score $\geq$ 80, and Variable Response Inconsistency $\geq$ 80)(Butcher & Williams, 1992). Three test protocols in the unscreened group were considered invalid and were eliminated from MMPI-2 comparisons. Demographic comparisons within this sample of 130 participants yielded the same results as before.

Participants' job performance was assessed via a questionnaire distributed to two of their direct supervisors (primarily persons occupying detachment commander and superintendent positions). The questionnaire was developed from a separate job analysis performed by detachment commanders and superintendents in 1994. In connection with another study, they had generated and defined ten skills considered key to successful agent performance: perception, decision making, decisiveness, organizing and planning, adaptability, interpersonal, control and follow-up, coaching, delegation, and communications. These skills were designed to be easily observable and distinct (M. A. Cooper, personal communication, February 28, 1994). The

42

questionnaire results used in the current study linked the ten skills to typical job examples, and behaviorally anchored them to a 5-point rating scale. The questionnaire also requested an overall evaluation of agent performance, also on a 5-point scale, as well as information about the number of positive and negative distinctions achieved. Positive distinctions included: letters of appreciation, designations as "agent of the quarter or year," awards and commendations, and distinguished graduate status (top 10% or better in training courses). Negative distinctions included: counselings, letters of reprimand, extended probations (lengthening of the mandatory one-year probation which follows academy training when program requirements are not met), citizen grievances, and administrative or congressional inquiries (evaluations of complaints against agents sent through headquarters or congressional delegates). Since inquiries, per se, are not inherently negative, they were counted against an individual only if they resulted in adverse outcome. Finally, the questionnaire requested demographic information relating to participants (age, rank, and academy graduation date) and two items relating to their supervisors (position and length of time having observed participant)(see Appendix 2 for questionnaire format).

Several steps were taken to minimize common rating errors due to indifference, prejudice, the halo effect, leniency, and error of central tendency. Supervisors were told that they had a unique opportunity to help examine whether using the MMPI-2 to screen applicants contributed over and above

43

other screening processes, and that they would save time and money if MMPI-2 characteristics were revealed which better predicted successful or unsuccessful agent performance. They were also assured that there was absolutely no linkage between their ratings and traditional performance reports or any other personnel decisions. Therefore, they were encouraged to give their most accurate assessment, free from any constraints about inflation or deflation. Finally, they were guaranteed all evaluations would remain strictly confidential. Procedures did not require the supervisors to be identified, as the questionnaires were distributed via unit addresses and supervisors put no personal information on them. They were told that all names and identifying features of the agents whose performance they rated would be removed once the data became linked to MMPI-2 scores. They were also assured that only group data would be reported in research results.

What constitutes successful law enforcement performance and how best to measure it remains highly debatable. The accuracy, validity, or superiority of supervisory ratings, behaviorally anchored rating scales, or other performance measures--subjective or objective--depends on purpose and situational constraints (Sulsky and Balzer, 1988). The performance measures in this study were chosen based on an analysis of the federal agency's organizational culture, interpersonal and job skills required, and previous performance studies.

In the first comparison of this study, MMPI-2 protocols from the unscreened group were contrasted with test profiles from the screened group to

determine whether they differed significantly. Given the limited sample size, it was not possible to compare all of the basic, supplementary, and content scales. Accordingly, this study used only scales (or combinations thereof) which appeared to have some empirical support for discriminating successful performance among law enforcement samples. These included three validity scales (L, F, and K), the ten clinical scales (Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma, and Si), the mean elevation for the ten clinical scales, and an "immaturity index" (raw scores of L, Pd, and Ma added together).

In addition, Ego Strength (Es) - K or "healthy defensiveness" was compared among the two groups. Although previous police officer research involved the original Es scale on the MMPI, and this scale was shortened considerably on the MMPI-2, Schuldberg (1992) reported that the revised Es scale compared favorably with the original on internal consistency. Additionally, previous normative studies involving both the MMPI and MMPI-2 conducted at the federal agency from whom the participants were drawn showed Es was high among successful agents (M. S. Roman, personal communication, 1990).

In the second comparison in this study, the job performance of the unscreened group was contrasted with the performance of the screened group. This actually involved various subsets of the total sample and a series of comparisons, as a substantial number of participants having MMPI-2 results on file did not have two performance questionnaires returned. The screened group had a 94% overall response rate. Of those, 81% had two raters and 19% had

45

one rater. The unscreened group had a 74% overall response rate. Of those, 81% had two raters and 19% had one rater. Additional tests were also needed because two subsets were created to more nearly control for experience. Thus, there were four comparisons between the unscreened and screened groups: (1) those having two raters--"2-rater" subset, $\underline{n}$ = 82, (2) those having at least one rater--"1-rater" subset, $\underline{n}$ = 102, (3) agents hired during 1995 and 1996 having two raters--"2-match" subset, $\underline{n}$ = 26, and (4) agents hired during 1995 and 1996 having at least one rater--"1-match" subset, $\underline{n}$ = 32.

As with the original total sample, there were no significant differences in the means for the screened and unscreened groups in any of the comparisons with respect to sex, race/ethnicity, education, and rank. In the 1-rater and 2-rater subsets, the screened groups were again slightly younger than the unscreened groups (both by about two years), although these differences were insignificant in practical terms. In both matched subsets, age no longer differed statistically between the screened and unscreened agents. Concerning experience, the screened groups in the 1-rater and 2-rater subsets had about one year's less experience than the unscreened groups, similar to the total group. In both matched groups, the screened groups had about three fewer months experience (about 12 months) than the unscreened groups (about 15 months) due to differing academy graduation dates. Although reaching statistical significance ($\underline{p}$ = .001 and .003 for the matched, 1- and 2-rater

subsets, respectively), the practical impact of such differences on performance ratings was theoretically minimal.

The supervisors' performance scores were averaged (for agents having two raters), because a large disparity did not exist among raters. For the screened group, exact agreement between the two raters across all ten performance subtests averaged 48 percent. Raters differing by one point averaged 46 percent, and by two points, 6 percent. None differed by more than two points. Percent agreement for the overall performance rating was somewhat higher: exact--61%, one point--31%, and two points--8%. For the unscreened group, exact agreement across all ten subtests averaged 49 percent. Raters differing by one point averaged 41 percent, and by two points, 9 percent. One percent differed by three points. This disagreement was troublesome, even though these ratings represented only one performance scale each for five different individuals. Four of the five, however, had either reprimands or very low scores on the other performance scales, so perhaps they were harder to rate. Percent agreement for the overall performance rating was somewhat higher: exact--58%, one-point--39%, and two-points--3%.

In addition to the predominately similar ratings given by raters, the supervision time was similar overall. The detachment commanders (or equivalents) had observed the participants from 2 to 33 months with the average being 13.26 months. Length of observation time did not statistically differ between the screened and unscreened groups. Of the subset of the total

47

sample who had only one rater, 75 percent of the screened and 76 percent of the unscreened group received scores from the detachment commander. The superintendents (or equivalents) had observed the participants from 2 to 36 months with the average being 13.93 months. Although similar to the numbers for detachment commanders, the length of observation time statistically differed among these raters. The screened group had about 4 fewer months ($\underline{M}$ = 10.50, $\underline{SD}$ = 4.62) than the unscreened group ($\underline{M}$ = 14.60, $\underline{SD}$ = 7.38), $\underline{t}$ (27.86) = -2.71, $\underline{p}$ = .011.

Although the ten performance skills and overall performance rating were designed to be separate, Cronbach's alpha coefficients were computed to determine if ratings could be collapsed to maximize power by reducing the number of statistically independent entities. The ten performance skills were highly related ($\underline{n}$ = 102, alpha = .938), as were the ten skills and the overall rating ($\underline{n}$ = 102, alpha = .949). Analyzed in a comparable way, the correlation coefficients for all 11 performance factors ranged from .350 (organization and interpersonal) to .779 (perception and decision making) with all $\underline{p}$ values being less than .0001. Accordingly, the mean of these 11 performance ratings was considered one performance criteria.

Consideration was given to whether the positive and negative performance criteria should be dichotomized (e.g., agents with a score of 1 or more contrasted with those who received zero scores), to prevent a few agents with positive scores from skewing the results. An organizational analysis had

suggested that the vast majority of agents might have few or no such distinctions. However, this did not prove to be the case. Seventy-five percent of the total participants earned positive distinctions, with the number ranging from zero to six ($\underline{M} = 1.84$, $\underline{SD} = 1.64$). Twenty-nine percent accumulated negative distinctions, with the number ranging from zero to four ($\underline{M} = .47$, $\underline{SD} = .86$). Positive distinctions were normally distributed and negative distinctions were only moderately skewed in a positive direction. Analyses were thus accomplished using actual values.

In the third comparison of this study, correlations between the MMPI-2 protocols and job performance of both groups were computed to determine which scales, if any, predicted job criteria. It was planned to compute correlations in either of two ways. If previous protocol and performance comparisons failed to yield significant differences, the unscreened and screened groups would be combined for greater predictive power. If, on the other hand, significant differences had resulted, the MMPI-2 profiles and job performance would be analyzed by group. Conceivably, it was possible for some statistical differences to arise which might not be of practical importance. In this eventuality, the correlations would also be computed on the total sample.

Concerning descriptive statistics, it was hypothesized a very large percentage of profiles (90% or greater) would reveal MMPI-2 scores in the normal range (T scores less than 65 on clinical scales). It was also expected that the most common validity profile would be the "inverted caret." Regarding

49

the first comparison, it was postulated the unscreened group would show greater MMPI-2 profile variability than the screened group, in terms of having a larger range of scale scores and larger standard deviations. It was also predicted that the unscreened group would tend to have higher mean scores than the screened group. However, it was unclear whether scale means would differ statistically. In respect to the second comparison, it was theorized the unscreened group(s) would not differ statistically from the screened group(s) on job performance, due to the samples being restricted by rigorous non-psychological screening. Finally, as to the third comparison, it was hypothesized that less successful agents would have more extreme L scores; higher F, 1 (Hs), and 9 (Ma) scores; higher overall distress; lower Es-K scores; and higher "immaturity index" scores as compared to more successful agents. Again, although these profile trends were expected, it was unclear they would reach statistical significance.

RESULTS

All descriptive and inferential analyses were accomplished using The
Statistical Package for the Social Sciences 6.0 (SPSS, Inc. 1993). Prior to
statistical testing, all data were screened for accuracy of data entry, missing
values, and fit between their distributions and the assumptions of univariate and
multivariate analysis.

Regarding MMPI-2 data, there were no differences between computer
scoring and manual replotting of a random sampling of protocols, and there were
no missing values. Three protocols from the unscreened group, however, were
eliminated for having nonresponsive or inconsistent response patterns, leaving
130 cases for analysis. Many MMPI-2 scale variables approximated normal
distributions. There was one exception in the screened group--scale 5 (Mf),
which had severe kurtosis. In the unscreened group, scales F and 0 (Si) had
severe positive kurtosis, 5 (Mf) and 8 (Sc) had moderate kurtosis, and 0 (Si) had
moderate skewness. The excepted distributions were not illogical, however, in
the sense that law enforcement personnel tend to cluster in lower T score
ranges. In any event, transforming the MMPI-2 data was not desired as it would
have led to uninterpretable values.

Similarly, some of the scales in the unscreened group contained outliers, if defined in the statistical sense of being more than three standard deviations from the mean: one each for scales F, D, Pd, Sc, Ma, and Si; and two for scale Mf. However, these eight points were considered meaningful rather than sources of error. For example, one of the 113 unscreened protocols had a T score of 70 on Scale 8 (Sc). Although in the clinical range, such a score is not invalid nor was it unreliable in the sense of being unknown in the larger population from which this sample was drawn. Although it was decided to retain such points, analyses were also performed without them in instances where their influence may have been questioned.

Descriptive Analyses

As hypothesized, the vast majority of MMPI-2 profiles (screened - 94%, unscreened - 93%) were in the normal range, and the inverted caret was the most common validity configuration in both groups. If defined loosely as those with scales L and K higher by 5 or more points than scale F, the percentages were: screened group = 94%, unscreened group = 90%. If strictly defined as L and $K \geq 60$ and $F \leq 50$, the figures were: screened = 24%, unscreened = 40%.

Using the mean profile for each group, the two-point code for screened agents was 9-4; for unscreened agents it was 4-9. Individuals who moderately elevate scale 9 (Ma) as in these profiles are described as energetic, uninhibited, extroverted, and talkative, whereas those who moderately elevate scale 4 (Pd)

tend towards confident, assertive, or interpersonally manipulative behavior and have definite opinions about right and wrong (Butcher & Williams, 1992).

The means, standard deviations, and ranges were computed and plotted for 16 MMPI-2 variables: three validity scales, ten clinical scales, "healthy defensiveness" (Es-K), "immaturity" (raw scores of L, Pd, and Ma added together), and "overall distress" (average of clinical scales). Table 1 contains the results and Figure 1 depicts the means of the traditional scales in typical MMPI-2 profile form.

Table 1

MMPI-2 Scale Means and Standard Deviations for Comparison Groups

| | Group | | | | | | | |
| | Screened (n = 17) | | | Unscreened (n = 113) | | | | |
Scale	M	SD	Range	M	SD	Range	t	p^*
L	56.94	7.72	43 - 70	59.25	9.75	35 - 83	-.93	.353
F	40.41	3.55	36 - 48	41.25	4.56	36 - 67	-.72	.471
K	58.00	6.83	47 - 68	61.32	7.02	41 - 74	-1.82	.071
1 (Hs)	46.00	4.62	39 - 57	47.83	6.21	35 - 66	-1.17	.245
2 (D)	44.00	4.17	36 - 54	44.44	6.12	30 - 68	-.29	.774

Table 1--continued

Scale	Screened (n = 17)			Unscreened (n = 113)			t	p*
	M	SD	Range	M	SD	Range		
3 (Hy)	45.88	5.52	38 - 57	47.91	6.16	31 - 64	-1.28	.202
4 (Pd)	47.65	6.17	39 - 63	49.62	6.30	34 - 69	-1.21	.230
5 (Mf)	42.35	12.28	30 - 79	42.46	9.65	30 - 77	-.04	.967
6 (Pa)	42.29	7.28	31 - 59	45.86	7.17	32 - 64	-1.91	.059
7 (Pt)	43.77	3.13	39 - 49	46.04	6.40	33 - 83	-1.43	.154
8 (Sc)	42.47	3.63	36 - 49	46.19	5.58	33 - 70	-2.66	.009
9 (Ma)	49.35	7.19	38 - 62	48.23	6.01	38 - 69	.70	.485
0 (Si)	42.00	6.00	35 - 57	41.54	7.36	30 - 84	.25	.806
Mean[a]	44.58	2.51	40 - 49[a]	46.01	3.67	37 - 63[a]	-1.56	.122
Es-K[b]	1.94	8.56	⁻12 - ⁺16	-.78	7.53	⁻21 - ⁺24	1.36	.175
Immat[c]	35.12	4.50	28 - 44	35.20	4.11	27 - 47	-.08	.937

Note. All values, excluding Es-K and Immaturity Index scores, represent T scores. Degrees of freedom = 128 for all analyses.

[a]Mean = the average of the ten clinical scales; range values are rounded to whole numbers for similarity to other columns. [b]Es-K = Ego Strength minus K or "healthy defensiveness." [c]Immat = Immaturity Index or the sum of raw scores for scales L, Pd, and M.

*p values are two-tailed.
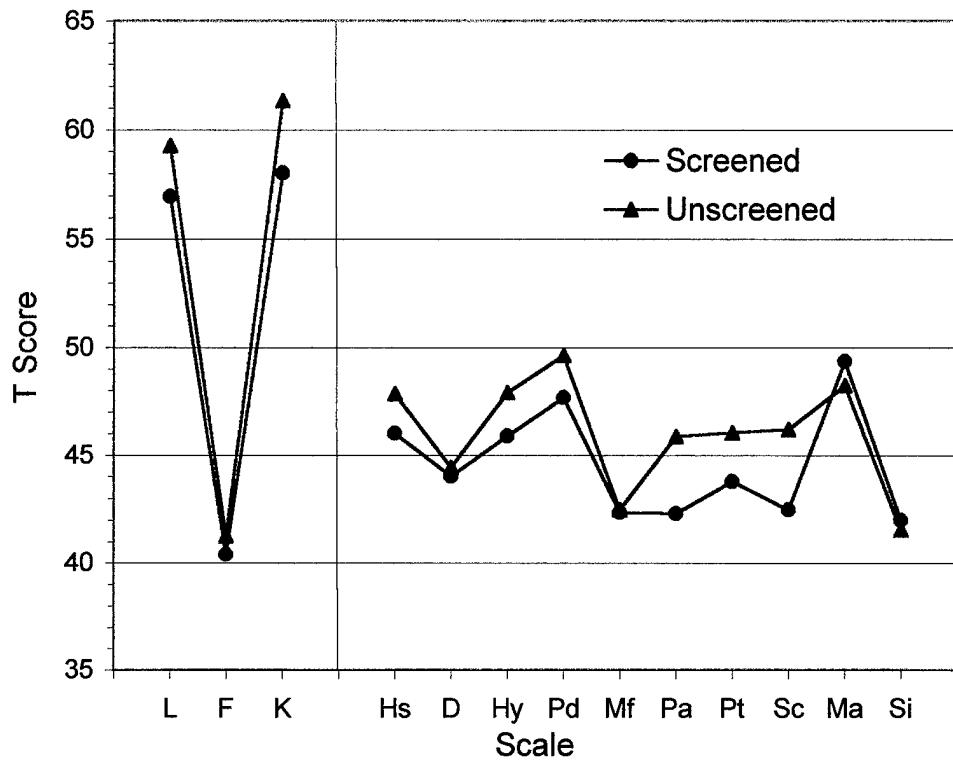
Figure 1. Mean MMPI-2 profiles for screened (n = 17) and unscreened (n = 113) groups.

Also tabulated were the frequencies of elevated scores (for current purposes, T scores falling in the ranges of 50-54, 55-59, 60-64, and above) for agents in each group (see Table 2). Given the unequal groups, percentages were given rather than numbers to facilitate comparison.

Comparison #1

From these two tables, several trends can be noted. The unscreened group had higher mean scores than the screened group on all validity scales, eight of the ten clinical scales, the mean of the ten clinical scales, and the

Table 2

Percentage of Elevated[a] MMPI-2 Scores for Comparison Groups

	T Scores							
	50 - 54		55 - 59		60 - 64		65 +	
Scale	Scr[b]	Uns[c]	Scr	Uns	Scr	Uns	Scr	Uns
L	11.8	20.4	17.6	15.9	23.5	16.8	23.6	34.5
F	0	3.5	0	0	0	0	0	.9
K	13.6	6.2	11.8	17.7	35.3	32.7	11.8	35.3
1 (Hs)	11.8	24.8	5.9	9.7	0	.9	0	.9
2 (D)	11.8	17.7	0	2.7	0	.9	0	.9
3 (Hy)	29.4	38.1	5.9	7.1	0	2.7	0	0
4 (Pd)	29.4	36.3	0	14.2	5.9	2.7	0	.9
5 (Mf)	0	7.1	0	3.5	5.9	2.7	5.9	3.5
6 (Pa)	5.9	15.0	5.9	7.1	0	3.5	0	0
7 (Pt)	0	14.2	0	7.1	0	0	0	.9
8 (Sc)	0	12.4	0	5.3	0	.9	0	.9
9 (Ma)	23.6	23.9	17.7	9.7	5.9	1.8	0	1.8
0 (Si)	5.9	4.4	5.9	1.8	0	1.8	0	.9
Mean[d]	0	10.6	0	0	0	.9	0	0

Note. All values are percentages.

56

Table 2--continued

^{a}Normally, a scale isn't considered "clinically" elevated until it reaches a

T score of 65 or greater. However, studies using the MMPI-2 with law

enforcement populations recommend lower cutoffs. ^{b}Scr = screened group,

$\underline{n}$ = 17. ^{c}Uns = unscreened group, $\underline{n}$ = 113. ^{d}Mean = the average of ten clinical

scales.

immaturity index. The unscreened group also had a smaller value of Es-K than

did the screened group. Further, the unscreened group had greater variability

than the screened group, in terms of a larger range of scores on every scale

except Mf and larger standard deviations on all validity and clinical scales except

Mf, Pa, and Ma. This was despite the fact that Levene's tests revealed the two

groups came from populations with equal variances. Finally, the unscreened

group had a higher frequency of elevated scale scores than did the screened

group, as seen by the higher percentages in 38 of the 56 T score range

comparisons (16 variables x 4 ranges). Such profile trends were as predicted.

To determine whether these trends were statistically different, an

independent, two-tailed t-test was performed on each of the 16 MMPI-2

variables. In reality, MMPI-2 variables are not independent because some

individual items contribute to more than one scale (even though they are

sometimes coded in opposite directions). Also, protected testing or a

simultaneous inference procedure would normally be used with multiple tests to

correct for family-wise error. Given the mixed literature results, however, the first

analysis was exploratory in nature. Thus, independent tests without family-wise

correction were employed to more likely detect any differences which existed in

this limited sample size.

In addition to means and standard deviations, Table 1 lists the results of

this significance testing. Only scale Sc reached statistical significance, although

Pa nearly did so. Listed in descending order, the remaining scales were: K,

Mean, Pt, Es-K, Hy, Pd, Hs, L, F, Ma, D, Si, Immaturity Index, and Mf. As Sc had

been one of the scales which contained an outlier, the t-test was recomputed

after deleting that case. The unscreened group remained statistically higher on

Sc than the screened group, t (127) = -2.72, p = .007.

Subsequently, a multivariate analysis of variance (MANOVA) was

computed to protect against inflated Type I error due to multiple tests of

correlated variables as well as to improve the chances of discovering any group

differences which might have been apparent only in combinations of variables.

There were 14 MMPI-2 variables in this planned comparison; Mean and

Immaturity Index were excluded because they were composites of other

variables and would have caused problems of singularity.

With the use of the Wilks' criterion, the unscreened group did not

statistically differ from the screened group, F (14, 115) = 1.046, p = .414. Sc

accounted for 5.25% of explained variance, Pa for 2.77%, and K for 2.53%. All

others were less than 2.00%, with the lowest--Mf--accounting for about .001%.

58

The explained variance across all MMPI-2 variables (1 - Wilks' lambda) was 11.30%.

In unplanned comparisons, multivariate models with subsets of MMPI-2 variables were analyzed. Using only Sc and Pa, there was an overall effect for group with $\underline{F}$ (2, 127) = 4.065, $\underline{p}$ = .019. Adding a third variable--K, the two groups again differed statistically, $\underline{F}$ (3, 126) = 2.829, $\underline{p}$ = .041. There was still an overall effect for group adding yet a fourth variable--Pt with $\underline{F}$ (4, 125) = 2.480, $\underline{p}$ = .047. However, no other variables could be contributed to the model while still attaining significance.

Comparison #2

As described in the Methods section, the performance data contained a substantial number of missing values. Two strategies were used to deal with this. First, it was deemed justifiable to combine agents having only one rater with those having two (the "1-rater subset," $\underline{n}$ = 102), given the noted similarities among raters. Skills and overall performance ratings were then averaged. Second, the results of this group were then compared with another group which included only agents having two raters (the "2-rater" subset, $\underline{n}$ = 82), to see whether they differed. Additional groups were also created to more nearly control for experience. Agents hired during 1995 and 1996 having at least one rater (the "1-match" subset, $\underline{n}$ = 32) were analyzed, as were agents hired during those years having two raters (the "2-match" subset, $\underline{n}$ = 26).

There were three performance variables: the mean of the ten performance skills and the overall rating (labeled "Rating"), number of positive distinctions ("Pos"), and number of negative distinctions ("Neg"). As described in the Methods section, the justification for creating the "Rating" variable was provided by Cronbach's alpha coefficients which revealed its component variables were highly correlated.

For the 1-rater subset, all variables approximated normal distributions in both the screened and unscreened groups, with the exception of "Neg" which had moderate positive kurtosis in the unscreened group. Such a distribution was not illogical, however, given the small range of negative distinctions (0 to 4) and the predominance of agents having zero or one values.

If defined in the statistical sense of being more than three standard deviations from the mean, the screened group contained no outliers on any of the variables whereas the unscreened group had one "Neg" outlier. However, this point was considered meaningful rather than a source of error. The individual had four discrete events which fell within the definition of the variable, and this number was not unreliable in the sense of being unknown in the larger population from which this sample was drawn. It was decided to retain this point.

Exploring the data yielded very similar results for the other three subsets. Only exceptions to normality of distribution or outliers will be noted, although the same comments made for the first subset would apply here as well. For the 2-rater subset, the unscreened group had moderate positive kurtosis for "Rating,"

and one outlier for "Neg." Neither the 1-match nor 2-match subsets had outliers, although the latter had moderate skewness and kurtosis in the unscreened group for "Neg."

Table 3 lists the means; standard deviations; and results of independent, two-tailed, t-testing for the three performance variables in the four subsets. Several trends can be seen. Unscreened agents had more distinctions, both positive and negative, than did screened agents for the larger two subsets (1- and 2-rater). Only the negative distinctions reached statistical significance, however (p = .021 and .026, respectively). In the two smaller subsets (1- and 2-match), these trends were eliminated. For all four subsets, unscreened and screened agents obtained approximately equivalent performance ratings.

Table 3

Means and Standard Deviations, by Comparison Group, for Four Performance Subsets

| | Group | | | | | | |
| | Screened | | Unscreened | | | | |
Subset	$\underline{M}$	$\underline{SD}$	$\underline{M}$	$\underline{SD}$	$\underline{df}$	$\underline{t}$	$\underline{p}$
1-rater	$\underline{n}$ = 16		$\underline{n}$ = 86				
Pos	1.44	1.15	1.92	1.71	28.93[a]	-1.41	.170
Neg	.19	.40	.52	.92	49.65	-2.38	.021
Rating	3.91	.66	.96	.69	100	-.24	.811

Table 3--continued

Subset	Screened		Unscreened				
	M	SD	M	SD	df	t	p
2-rater	n = 13		n = 69				
Pos	1.38	1.26	1.90	1.58	80	-1.10	.273
Neg	.23	.44	.62	.99	39.47	-2.31	.026
Rating	4.03	.67	3.97	.67	80	.31	.756
1-match	n = 16		n = 16				
Pos	1.44	1.15	1.15	1.38	30	.15	.879
Neg	.19	.40	.40	.25	30	-.42	.681
Rating	3.91	.66	.66	3.97	30	-.24	.808
2-match	n = 13		n = 13				
Pos	1.38	1.26	1.54	1.20	24	-.32	.753
Neg	.23	.44	.15	.38	24	.48	.635
Rating	4.03	.67	4.05	.51	24	-.06	.953

Note. Pos = number of positive distinctions. Neg = number of negative distinctions. Rating = the mean of the ten performance skills and overall performance rating. 1-rater = subset having at least one rater. 2-rater = subset having two raters. 1-match = subset hired in 1995/96 having at least one rater. 2-match = subset hired in 1995/96 having two raters.

[a]Degrees of freedom with fractions represent corrections for unequal variance, per Levene's Test.

Although this exploratory analysis considered the t-tests individually, a more accurate test of significance would incorporate protection against the inflation of error rate inherent in multiple t-testing. Thus, two correction procedures were employed. First, simultaneous inference was accomplished by establishing the family-wise alpha level at .05, which then established the corresponding test-wise error rate at .017 (alpha divided by three). Using this method, negative distinctions among unscreened and screened agents in the 1- and 2-rater subsets no longer reached statistical significance.

Subsequently, protected testing was accomplished via MANOVAs. Using the Wilks' criterion, none of the subsets showed an overall effect for group. For the 1-rater subset, F (3, 98) = 1.08, p = .360. The 2-rater subset had F (3, 78) = .976, p = .408. In the 1-match subset, F (3, 28) = .10, p = .959. The 2-match subset had F (3, 22) = .11, p = .952. Because the global null hypothesis could not be rejected, no follow-up tests for specific differences were conducted. Across the four subsets, the percent of variance explained by "Pos" ranged from 1.50% to .08%, by "Neg" from 2.40% to .57%, and by "Rating" from .20% to .02%. The explained variance accounted for by all performance variables (1 - Wilks' lambda) ranged from 3.62% in the 2-rater subset to 1.07% in the 1-match subset).

Albeit insignificant in protected testing, the trend for unscreened agents in the larger subsets to have more negative distinctions than screened agents, as well as the trend for larger subsets to have a couple of more percentage points

in total variance explained by performance measures, appeared to be due to experience. Recall that the larger subsets differed by about 12 months in experience while the smaller ones were more nearly equal. However, this result may have been confounded with the smaller sample size of the experience-matched subsets.

Accordingly, two more MANOVAs were computed for the 1- and 2-rater subsets using experience as a covariate. Using the Wilks' criterion, in the 1-rater subset the overall p value changed from .360 (without experience as a covariate) to .875 (with it), F (3, 97) = .230. In the 2-rater" subset the p value changed from .408 to .881, F (3, 77) = .222.

The failure to find statistically significant differences between the screened and unscreened agents on job performance was as hypothesized. The underlying reasons for this, however, may have involved not only the presumed contribution of restricted samples but also the confounding of experience and two of the performance measures.

Comparison #3

Given that the screened and unscreened groups did not differ on either MMPI-2 or performance analyses, the groups were combined to maximize sample size and thus power. In the final comparison, correlation coefficients were computed to investigate any predictive links between MMPI-2 scales and job performance criteria. Two subsets were used: 1-rater -- because it was the largest subset, and 2-rater -- because it was also a large subset but arguably

64

may have been more reliable in that all agents were rated by two supervisors. Examination of the scatterplots revealed linearity was satisfactory. Although the small effect sizes limit the degree of linearity observed, no other shape better accounted for the relationships.

Table 4 gives the Pearson Product Moment correlations and associated p values for the MMPI-2 scales in relation to performance variables.

Table 4

Correlations for MMPI-2 Scales in Relation to Performance Variables

Scale	Pos		Neg		Rating	
	r	p	r	p	r	p
1-Rater Subset (n = 102)						
L	-.012	.905	.144	.150	-.123	.218
F	-.030	.767	.070	.484	-.107	.283
K	.103	.303	.148	.138	.216	.029*
1 (Hs)	.145	.146	.236	.017*	.112	.261
2 (D)	.094	.345	.106	.290	-.078	.434
3 (Hy)	.114	.253	.291	.003**	.060	.550
4 (Pd)	.189	.057	.229	.021*	.147	.139
5 (Mf)	.021	.834	.030	.767	-.127	.204
6 (Pa)	.040	.690	.233	.019*	-.008	.933
7 (Pt)	.091	.366	.025	.807	.175	.078
8 (Sc)	.101	.313	.072	.413	.209	.035*
9 (Ma)	-.070	.485	-.083	.408	-.004	.969
0 (Si)	.027	.787	.008	.934	-.055	.581
Mean	.127	.205	.201	.043*	.054	.588
Es-K	-.130	.193	-.106	.288	-.123	.217
Immat	.023	.816	.103	.304	-.062	.536

Table 4--continued

Scale	Pos		Neg		Rating	
	r	p	r	p	r	p
			2-Rater Subset (n = 82)			
L	-.058	.608	.113	.312	-.132	.238
F	.006	.955	.061	.586	-.176	.115
K	.024	.833	.098	.381	.235	.034*
1 (Hs)	.122	.275	.222	.045*	.092	.409
2 (D)	.099	.376	.090	.420	-.087	.435
3 (Hy)	.116	.299	.296	.007**	.076	.499
4 (Pd)	.204	.066	.184	.098	.121	.279
5 (Mf)	.033	.766	.041	.717	-.170	.126
6 (Pa)	.016	.885	.213	.054	-.064	.568
7 (Pt)	.080	.476	-.022	.847	.124	.268
8 (Sc)	.116	.300	.030	.792	.142	.204
9 (Ma)	-.063	.573	-.081	.467	-.093	.406
0 (Si)	.114	.310	-.010	.932	-.090	.423
Mean[a]	.140	.210	.168	.131	-.010	.926
Es-K[b]	-.090	.424	-.066	.557	-.128	.252
Immat[c]	.054	.630	.085	.449	-.155	.165

Note. Pos = number of positive distinctions. Neg = number of negative distinctions. Rating = the mean of the ten performance skills and overall performance rating. 1-rater = subset having at least one rater. 2-rater = subset having two raters.

[a]Mean = the average of the ten clinical scales. [b]Es-K = Ego Strength minus K or "healthy defensiveness." [c]Immat = Immaturity Index or the sum of raw scores for scales L, Pd, and M.

*p < .05. **p < .01. All values are two-tailed.

Looking at positive distinctions, in both the 1- and 2-rater subsets there were no MMPI-2 scales which significantly correlated with "Pos," although Pd was close. Concerning negative distinctions, in the 1-rater subset there were five MMPI-2 scales which significantly correlated with "Neg" (in order from most to least significant): Hy, Hs, Pa, Pd, and Mean. For the 2-rater subset, two of these scales attained significance: Hy and Hs (in that order). Pa was nearly significant.

Regarding the combined performance rating, in the 1-rater subset both K and Sc (in descending order) significantly correlated with "Rating," whereas in the 2-rater subset, only K did so. Each of the noted correlations (for all three performance variables and both subsets) were positive; that is, the higher the MMPI-2 T score, the higher the positive and negative distinctions or the higher the combined rating.

These results were largely contrary to predictions. Less successful agents did not have more extreme L scores; higher F and Ma scores; lower Es-K scores; and higher "immaturity index" scores. They did have higher Hs scores, however, and higher "Mean" or overall distress (at least for one of the subsets). Finally, although it was not hypothesized, the result that more successful agents had higher K scores was consistent with cited literature results.

DISCUSSION

Psychologists and employers should adhere to key preemployment psychological screening guidelines. These include using psychological evaluation results as one component of the overall selection process, knowing how such results apply to law enforcement candidates, and being prepared to defend assessment practices and selection recommendations (Scogin & Beutler, 1985).

To this end, many researchers have examined the utility of the MMPI and MMPI-2 with regard to specific law enforcement populations and specific job selection questions. The current study looked at a sample of military federal investigators to determine whether particular MMPI-2 scales differentially predicted organizationally tailored performance criteria.

If the unscreened group had shown greater psychological profile variability than the screened group, as determined by the MANOVA, it would have suggested the MMPI-2 added information beyond that obtained through non-psychological screening measures. Given that the unscreened group did not significantly differ from the screened group on MMPI-2 scales, no specific conclusions can be drawn. Failing to reject the null hypothesis is always

difficult to explain given the multitude of possible reasons why any differences that exist may go unnoticed.

In this study, possibilities included small sample size (and thus, limited statistical power), prior attenuation of the unscreened group due to non-psychological screening, attenuation of the unscreened group due to study design, and lack of representativeness of the screened group. Study design may have reduced differences in that it was not possible to study the MMPI-2 scores of applicants who were no longer on the job (because of the inability to collect biographic data and performance measures). They may have voluntarily withdrawn or been removed from investigative duty due to psychological incompatibility with the demands of the career field. One descriptive study of 239 applicants found that 7% more of those with "unresolved--clarification required" MMPI test results self-eliminated from the organization than those without indicators of psychological concern (N. S. Hibler, personal communication, April 25, 1984).

Results of the more liberal testing (unprotected t-tests and exploratory MANOVAs) as well as descriptive statistical trends suggested that differences between the screened and unscreened groups may have existed. For example, the unscreened group had higher mean scores on the majority of scales, a greater frequency of elevated scale scores, and greater variability. Had the sample sizes been larger, and had more theoretical justification existed for extracting a smaller subset of MMPI-2 scales to examine in this particular

population, conclusions may have differed. Nevertheless, even if the MMPI-2 had added information beyond that obtained from non-psychological screening, it might not have been enough for practical significance. As often noted with other law enforcement samples, differences in T score means between groups were very small, usually only a few points.

The second exploration was whether the unscreened group differed significantly from the screened group on job performance, as determined by the family-wise t-tests or MANOVAs. Given that no differences resulted, it seems most likely that previous non-psychological screening substantially reduced the magnitude of predictor-criterion relationships, theoretical approach notwithstanding.

MANOVA results using experience as a covariate also suggested that experience was confounded with two of the performance variables. That is, more experienced agents tended to have greater positive and negative distinctions. Screened agents, having been hired later, apparently lacked sufficient time to accumulate distinctions. This differed from other research results which showed that the majority of police officers who received serious disciplinary action or were fired got into trouble during their first year on the job (Bartol, 1991; Hiatt & Hargrave, 1988a). All of the agents had at least one year's experience.

The third comparison involved exploring any predictive links between MMPI-2 protocols and job performance criteria. As determined by correlation

coefficients using the total sample from the larger subsets, no MMPI-2 scales were significantly linked to positive distinctions, five scales (Hy, Hs, Pa, Pd, and Mean) were related to negative distinctions, and two scales (K and Sc) were correlated with the combined performance rating. The practical significance of these results is limited, in that none of the correlations were of sufficient magnitude to be used as predictive measures by themselves.

With the results of these three comparisons, some might assert the MMPI-2's usefulness is questionable. Yet it might be more accurate to say its utility is limited, at present, but not fully explored. Whereas "screening in" successful performers is a difficult undertaking, "screening out" applicants with indicators of psychopathology is much easier. Although no one in this unscreened sample evidenced obvious psychological problems, past MMPI and MMPI-2 testing of other agent applicants has occasionally revealed serious emotional instability. In addition to the obvious ethical implications, screening out such applicants saved roughly $15,000 per agent in initial training costs.

Realistically, one would have to be cautious in generalizing the results to other agencies, even to federal agencies with similar performance criteria or to other selected-duty military populations (e.g., air- and space-crew, nuclear operators, White House and National Communications Agency communicators, etc.). This does not mean that confidence in psychological screening should be abandoned, however, since the constraints of this study would have likely underestimated existing differences.

The potential for MMPI-2 protocols to differentially predict job performance was suggested, but left untapped, in a study by Flynn, Sipes, Grosenbach, and Ellsworth (1994). They demonstrated that Air Force F-16 aviators could agree on who were top performers and what personal qualities were important for that distinction. They also found that the pattern of MMPI-2 scores was similar to that found in an earlier retrospective study comparing Army and Navy pilots' scores to older Air Force pilots' norms (which used the MMPI). Unfortunately, this study did not attempt to correlate testing profiles with top performer status.

The current study contributed useful knowledge despite statistically insignificant or unimportant results. It capitalized on a unique sample of agents permitted to work without benefit of psychological test results--a methodological ideal. It also systematically studied selection assumptions regarding the contribution of psychological screening. If differences had resulted, employment selection procedures could have changed. For example, the relatively quick and inexpensive MMPI-2 assessment could have occurred earlier in the selection or performance prediction process, thereby saving time and resources. As it stands, continued systematic investigation is needed before employment selection assumptions can be confirmed or refuted, and procedures changed accordingly.

APPENDIX 1

Table 5

MMPI-2 Scale Designations, Descriptors, and Relevance to Police Officer

Performance

MMPI-2 Scale	Description	Relevance
L Lie	Measures the tendency to distort responses by claiming unrealistically favorable view of moral character and psychological adjustment. Items obvious in content.	Success: moderate scores. Failure: very low/high scores.
F Infrequency (Faking)	Measures symptom exaggeration due to faking, severe psychopathology, disorientation, or malingering. Items endorsed by < 10% of normal adult sample.	Failure: high scores.
K Correction (Defensiveness)	Measures test defensiveness and corrects for the tendency to deny problems. Items less obvious in content. Also reflects level of coping resources.	Success: high scores.
1 (Hs) Hypochondriasis	Measures abnormal, psychoneurotic concern over bodily health. Also reflects self-centeredness, whininess.	Failure: high scores.
2 (D) Depression	Measures negative frame of mind, poor morale, lack of hope in future, dissatisfaction with life, and low mood.	Mentioned less often.

Table 5--continued

MMPI-2 Scale	Description	Relevance
3 (Hy) Hysteria	Measures three clusters: psychological denial, social facility/assertiveness, and manifestation of vague somatic complaints. Job applicants can elevate scale by endorsing items in first two clusters.	Mixed findings; higher eleva-tions for both success and failure.
4 (Pd) Psychopathic Deviate	Measures antisocial tendencies including family discord, authority problems, social or self alienation, and social confidence. Also reflects extroverted lifestyles.	Mixed; higher eleva-tions for both success and failure.
5 (Mf) Masculinity/ Femininity	Measures stereotypically masculine or feminine interests, values, and person-ality characteristics.	Mentioned less often.
6 (Pa) Paranoia	Measures suspiciousness, mistrust, rigid thinking, excessive interpersonal sensiti-vity, and externalization of blame.	Mixed; higher eleva-tions for both success and failure.
7 (Pt) Psychasthenia	Measures anxiousness, severe rumina-tions, and obsessive-compulsive features. Also reflects concentration difficulties, indecisiveness, and perfectionism.	Mentioned less often.
8 (Sc) Schizophrenia	Measures social and emotional alienation, lack of ego mastery, strange thoughts, and bizarre sensory experiences. Also reflects unconventional lifestyle, nonconformism.	Mentioned less often.

Table 5--continued

MMPI-2 Scale	Description	Relevance
9 (Ma) Hypomania	Measures tendency to act in euphoric, aggressive, and hyperactive ways. Also reflects amorality or guilefulness and talkative, energetic lifestyle.	Failure: high scores.
0 (Si) Social Introversion	Measures shyness, social avoidance, and self-other alienation (high scores) or social extroversion, gregariousness (low scores).	Mentioned less often.

Note. Descriptors were synthesized from Butcher and Williams (1992). Except as noted, they describe elevated scale scores. Their relevance to successful or unsuccessful police officer performance was synthesized from the literature review chapter. Notice that a high score on one dimension (e.g., high Ma for unsuccessful performance) does not necessarily indicate that all other scoring possibilities reflect the other dimension (e.g., moderate or low Ma reflects successful performance).

APPENDIX 2

PERFORMANCE QUESTIONNAIRE FOR USE WITH MMPI-2 STUDY

Individual to be Evaluated: _____   Age: _____

Date Individual Graduated from (Agency) Academy: _____   Rank: _____

Rater's Position: _____

Length of Time Rater has Observed Individual (Please list number of months): _____

Skills: Please read the definition of all ten skills first, and attempt to keep them separate. Consider specific observations or samples of behavior you have made. Then circle the most appropriate number for each skill.)

A. Perception - Identifies critical pieces of information and elements of a situation; interprets and evaluates their meaning in the context of available data. (Ex. Recognizes key facts when interviewing witness, reading staff summary.)

1	2	3	4	5
Misses major elements; often has faulty interpretation		IDs major elements but misses minor ones; usually interprets correctly		IDs major & minor elements; consistently interprets correctly

B. Decision Making - Determines logical courses of action in addressing problems and issues; comes to rational conclusions based on available supporting data. (Includes the factor of quality.) (Ex. Opens investigation or refers information elsewhere; requests expert assistance when personal cabalilities are exceeded.)

1	2	3	4	5
Makes illogical decisions; doesn't fit data		Usually makes logical decisions; may skip some data		Consistently makes logical decisions; uses all data

C. Decisiveness - Initiates action; demonstrates little hesitancy in making decisions. (Does not include quality factor. Involves responding quickly, withstanding challenges, using confident tone of voice and body behavior.) (Ex. Controls chaotic crime scene; handles unusual phone inquiries.)

1	2	3	4	5
Often stuck in indecision; not confident		Usually rises to challenges; may show some hesitancy		Rises to all challenges; appears confident

D. Organizing and Planning - Systematically structures activities; establishes priorities and strategies for accomplishing specific results. (Ex. Plans and orders investigative steps; identifies and fills training gaps; prioritizes suspenses.)

1	2	3	4	5
Acts haphazardly; strategies don't match task		Usually acts systematically; some strategies ineffective		Has well-established system; priorities & strategies well linked

E. Adaptability - Adjusts one's behavior or approaches according to varying situations and changing demands. (Includes ability to manage stress.) (Ex. Puts "dirty" source and "respected" authority equally at ease.)

1	2	3	4	5
Rigid; breaks down under pressure		Usually flexible, sometimes slow to see uniqueness; good under pressure		Flexible, uniquely creative; excels under pressure

F. Interpersonal - Behaves in a manner which reflects sensitivity to the needs, feelings, and capabilities of others; sensitive to political considerations; rejects requests or proposals without offending others; listens. (Ex. Finds a way to help Unit/CC while denying investigation; constructively criticizes colleague in private; calms scared victim.)

1	2	3	4	5
Insensitive; tactless; politically incorrect		Usually sensitive; tactful; politically aware		Highly sensitive; tactful; a political master

G. Control and Follow-Up - Monitors and measures work progress or performance; ensures previous commitments are adhered to. (Has a group activity connotation.) (Ex. Documents command action on cases; fixes inspection deficits.)

1	2	3	4	5
Does not measure performance; activities not completed		Usually measures progress; completes many activities		Always knows meaningful status; completes all activities

H. Coaching - Accomplishes objectives/results by guiding subordinates' (or others') activities. (Has a one-to-one connotation; may extend beyond immediate task.) (Ex. Helps agent conceptualize case, grow in confidence.)

1	2	3	4	5
Coaching absent; not motivating		Good coach and motivator; guides on current tasks		Superb coach and motivator; develops for future tasks

I. Delegation - Assigns work to subordinates (or parcels up work with others) consistent with their capabilities and experience. (Ex. Assigns program managers according to skills and allows them full authority and responsibility.)

1	2	3	4	5
Does not delegate or does so illogically		Usually delegates consistent with capabilities/experience		Delegates masterfully; maximizes capabilities/experience

J. Communications - Expresses oneself clearly through both oral and written means; effectively uses voice tone, inflection, eye contact, and gestures when speaking; effectively uses technical factors such as grammar, spelling, and punctuation when writing. (Ex. Briefs cases/program status clearly and concisely; writes well-organized letters/reports.)

1	2	3	4	5
Hard to follow; points unclear; many errors		Usually expresses points clearly; may be stronger orator than writer (or vice versa)		Skilled orator and writer; points expressed clearly and powerfully

Overall Evaluation of Agent Performance - Consider all aspects of job performance.

1	2	3	4	5
Very poor performer; needs improvement in multiple areas		Average performer; generally well-rounded		Exceptional performer; multi-talented

Number of Positive Performance Distinctions While an Agent (examples: letter of appreciation, agent of quarter/year, award/commendation, distinguished graduate): _____

Number of Negative Performance Distinctions While an Agent (examples: counseling, letter of reprimand, extended probation, citizen grievance, administrative/congressional inquiry): _____

Thank you for your help. Please keep confidential.

REFERENCES

Azen, S. P., Snibbe, H. M., & Montgomery, H. R. (1973). A longitudinal predictive study of success and performance of law enforcement officers. Journal of Applied Psychology, 57, (2), 190-192.

Barrick, M. R., & Mount, M. K. (1991). The Big-Five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-26.

Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big-Five personality dimensions and job performance. Journal of Applied Psychology, 78, (1), 111-118.

Bartol, C. R. (1991). Predictive validation of the MMPI for small-town police officers who fail. Professional Psychology: Research and Practice, 22, (2), 127-132.

Beutler, L.E., Storm, A., Kirkish, P., Scogin, F., & Gaines, J. A. (1985). Parameters in the prediction of police officer performance. Professional Psychology: Research and Practice, 16, (2), 324-335.

Blau, T. H. (1994). Psychological services for law enforcement. New York: John Wiley & Sons.

Blau, T. H., Super, J., & Brady, L. (1993). The MMPI good cop/bad cop profile in identifying dysfunctional law enforcement personnel. Journal of Police and Criminal Psychology, 9, (1), 2-4.

Borum, R., & Stock, H. V. (1993). Detection of deception in law enforcement applicants: A preliminary investigation. Law and Human Behavior, 17, (2), 157-166.

Burbeck. E., & Furnham, A. (1985). Police officer selection: A critical review of the literature. Journal of Police Science and Administration, 13, (1), 58-69.

Butcher, J. N., Graham, J. R., & Ben-Porath, Y. S. (1995). Methodological problems and issues in MMPI, MMPI-2, and MMPI-A research. Psychological

Assessment, 7, (3), 320-329.

Butcher, J. N., Jeffrey, T. B., Cayton, T. G., & Colligan, S. (1990). A study of active duty military personnel with the MMPI-2. Military Psychology, 2, (1), 47-61.

Butcher, J. N., & Williams, C. L. (1992). Essentials of MMPI-2 and MMPI-A interpretation. Minneapolis, MN: University of Minnesota Press.

Campos, L. P. (1989). Adverse impact, unfairness, and bias in the psychological screening of Hispanic peace officers. Hispanic Journal of Behavioral Sciences, 11, (2) 122-135.

Duckworth, J. C., & Anderson, W. P. (1995). MMPI & MMPI-2 interpretation manual for counselors and clinicians (4th ed.). Bristol, PA: Accelerated Development.

Fabricatore, J., Azen, S., Schoentgen, S., & Snibbe, H. (1978). Predicting performance of police officers using the Sixteen Personality Factor Questionnaire. American Journal of Community Psychology, 6, (1), 63-70.

Flynn, C. F., Sipes, W. E., Grosenbach, J. J., & Ellsworth, J. (1994). Top performer survey: Computerized psychological assessment in aircrew. Aviation, Space, and Environmental Medicine, 65, (5, Sect. 2, Suppl.), A39-A44.

Gottesman, J. I., (1975). The utility of the MMPI in assessing the personality patterns of urban police applicants. Dissertation Abstracts International, 36 (09), 4743B.

Greene, R. (1991). The MMPI-2/MMPI: An interpretive manual. Boston, MA: Allyn and Bacon.

Grossman. L. S., Haywood, T. W., Ostrov, E., Wasyliw, O., & Cavanaugh, J. L. (1990). Sensitivity of MMPI validity scales to motivational factors in psychological evaluations of police officers. Journal of Personality Assessment, 55, 549-561.

Hargrave, G. E. (1985). Using the MMPI and CPI to screen law enforcement applicants: A study of reliability and validity of clinicians' decisions. Journal of Police Science and Administration, 13, (3), 220-224.

Hargrave, G. E., & Hiatt, D. (1987). Law enforcement selection with the interview, MMPI, and CPI: A study of reliability and validity. Journal of Police Science and Administration, 15, 110-117.

Hargrave, G. E., Hiatt, D., Ogard, E. M., & Karr, C. (1994). Comparison of the MMPI and the MMPI-2 for a sample of peace officers. Psychological Assessment, 6, (1), 27-32.

Hiatt, D., & Hargrave, G. E. (1988a). MMPI profiles of problem peace officers. Journal of Personality Assessment, 52, (4), 722-731.

Hiatt, D., & Hargrave, G. E. (1988b). Predicting job performance problems with psychological screening. Journal of Police Science and Administration, 16, (2), 122-125.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions. American Psychologist, 51, (5), 469-477.

Hogan, R., & Kurtines, W. (1975). Personological correlates of police effectiveness. The Journal of Psychology, 91, 289-295.

Hollrah, J. L., Schlottmann, R. S., Scott, A. B., & Brunetti, D. G. (1995). Validity of the MMPI subtle items. Journal of Personality Assessment, 65, (2), 278-299.

Johnson, E. E. (1983). Psychological tests used in assessing a sample of police and fire fighter candidates. Journal of Police Science and Administration, 11, (4), 430-433.

Keiller, S. W., & Graham, K. R. (1993). The meaning of low scores on the MMPI-2 clinical scales of normal subjects. Journal of Personality Assessment, 61, 211-223.

Kornfeld, A. D. (1995). Police officer candidate MMPI-2 performance: Gender, ethnic, and normative factors. Journal of Clinical Psychology, 51, 4, 536-540.

Lester, D., Babcock, S. D., Cassisi, J. P., & Brunetta, M. (1980). Hiring despite the psychologist's objections: An evaluation of psychological evaluations of police officers. Criminal Justice and Behavior, 7, (1), 41-49.

Levy, R. J. (1967). Predicting police failures. Journal of Criminal Law, Criminology, and Police Science, 55, (2) 371-378.

Matarazzo, J. D., Allen, B. V., Saslow, G., & Wiens, A. N. (1964). Characteristics of successful policemen and firemen applicants. Journal of Applied Psychology, 48, (2), 123-133.

Merian, E M., Stefan, D., Schoenfeld, L. S., & Kobos, J. C. (1980). Screening of police applicants: A 5-item MMPI research index. Psychological Reports, 47, 155-158.

Mills, M. C., & Stratton, J. G. (1982). The MMPI and the prediction of police job performance. FBI Law Enforcement Bulletin, 51, 11-15.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. Journal of Applied Psychology, 78, (4), 679-703.

Rogers, T. B. (1995). The psychological testing enterprise: An introduction. Pacific Grove, CA: Brooks/Cole.

Saxe, S. J., & Reiser, M. (1976). A comparison of three police applicant groups using the MMPI. Journal of Police Science and Administration, 4, (4), 419-425.

Schoenfeld, L. S., Kobos, J. C., & Phinney, I. R. (1980). Screening police applicants: A study of reliability with the MMPI. Psychological Reports, 47, 419-425.

Schuldberg, D. (1992). Ego-strength revised: A comparison of the MMPI-2 and MMPI-1 versions of the Barron Ego Strength scale. Journal of Clinical Psychology, 48, (4), 500-505. (Note: Have not cited yet.)

Scogin, F. & Beutler, L. E. (1985). Psychological screening of law enforcement candidates. In P. A. Keller & L. G. Ritt (Eds.), Innovations in clinical practice: A source book (Vol. 5, pp. 317-330). Sarasota, FL: Professional Resource Exchange.

Shusman, E. J., Inwald, R. E., & Knatz, H. F. (1987). A cross-validation study of police recruit performance as predicted by the IPI and MMPI. Journal of Police Science and Administration, 15, (2), 162-169.

Statistical Package for the Social Sciences (Version 6.0) [Computer software]. Chicago: SPSS.

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73, (3), 497-506.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. <u>Personnel Psychology, 44,</u> 703-742.

Weed, N. C., Ben-Porath, Y.S., & Butcher, J. N. (1990). Failure of Weiner and Harmon Minnesota Multiphasic Personality Inventory (MMPI) subtle scales as personality descriptors and as validity indicators. <u>Psychological Assessment, 2,</u> (3), 281-285.

Wright, B. S., Doerner, W. G., & Speir, J. C. (1990). Pre-employment psychological testing as a predictor of police performance during an FTO program. <u>American Journal of Police, 9,</u> (4), 65-84.

BIOGRAPHICAL SKETCH

Ann P. Funk was born on March 16, 1960, in Redwood City, California. She attended the University of Kansas on a four-year Air Force scholarship, graduating in May 1981 with BA degrees in Psychology and Crime & Delinquency Studies. She also earned Distinguished Graduate status from the Air Force Reserve Officer Training Corps. Following graduation, she began a 16-year Air Force career, holding command, staff, operations, and maintenance positions in the communications and investigations career fields. She completed professional military education in residence, Squadron Officers' School in 1987 and Air Command and Staff College in 1995. In May 1990 she earned an MA degree in Criminal Justice from George Washington University, and in June 1995 began an Air Force sponsored doctoral program in clinical psychology at Florida State University. She has received numerous awards and decorations, including the AF Meritorious Service Medal, AF Commendation Medal, AF Achievement Medal, AF Organizational Excellence Award, AF Recognition Ribbon, and National Defense Service Medal. Currently, she holds the rank of Major (Lieutenant Colonel select). Ann is married and has one son.